Essays in Misallocation and Economic Development

by

Xican Xi

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2016 by the
Graduate Supervisory Committee:

Berthold Herrendorf, Co-Chair
Gustavo Ventura, Co-Chair
Todd Schoellman

ARIZONA STATE UNIVERSITY

May 2016

ABSTRACT

The dissertation consists of two essays in misallocation and development. In particular, the essays explore how government policies distort resource allocation across production units, and therefore affect aggregate economic and environmental outcomes.

The first chapter studies the aggregate consequences of misallocation in a firm dynamics model with multi-establishment firms. I calibrate my model to the US firm size distribution with respect to both the number of employees and the number of establishments, and use it to study distortions that are correlated with establishment size, or so-called size-dependent distortions to establishments, which are modeled as implicit output taxes. In contrast to previous studies, I find that size-dependent distortions are not more damaging to aggregate productivity and output than size-independent distortions, while the implicit tax revenue approximately summarizes the effects on aggregate output. I also use the model to compare the effects of size-dependent distortions to establishments and to firms, and find that they have different effects on firm size distribution, but have similar effects on aggregate output.

The second chapter studies the effects of product market frictions on firm size distribution and their implications for industrial pollution in China. Using a unique micro-level manufacturing census, I find that larger firms generate and emit less pollutants per unit of production. I also provide evidence suggesting the existence of size-dependent product market frictions that disproportionately affect larger firms. Using a model with firms heterogeneous in productivity and an endogenous choice of pollution treatment technology, I show that these frictions result in lower adoption rate of clean technology, higher pollution and lower aggregate output. I use the model to evaluate policies that eliminate size-dependent frictions, and those that increase environmental regulation. Quantitative results show that eliminating size-

dependent frictions increases output by 30%. Meanwhile, the fraction of firms using clean technology increases by 27% and aggregate pollution decreases by 20%. In contrast, a regulatory policy which increases the clean technology adoption rate by the same 27%, has no effect on aggregate output and leads to only 10% reduction in aggregate pollution.

To my parents and Haiyan.

ACKNOWLEDGMENTS

I owe a great debt of gratitude to my dissertation committee members: Berthold Herrendorf, Gustavo Ventura, and Todd Schoellman. I can't possibly thank my co-advisors Berthold and Gustavo enough for their selfless dedication and generous support. Each of them gave far more of their time and effort than I deserved, and encouraged me to push through the ups and downs of research through the years. This dissertation would have been impossible without their inputs. I would also like to extend my sincere thanks to Todd. He taught me many things, brought many insightful questions into my research, and was always available when I needed his help.

I must acknowledge the numerous other ASU faculty members who contributed to my graduate education and provided useful advice and suggestions on my research, both current and past, including Seung Ahn, Alexander Bick, Hector Chade, Madhav Chandrasekher, Manjira Datta, Domenico Ferraro, Amanda Friedenberg, Bart Hobijn, Roozbeh Hosseini, Natalia Kovrijnykh, Nick Kuminoff, David Lagakos, Edward C. Prescott, Natalia Ramondo, B Ravikumar, Kevin Reffett, Richard Rogerson, Edward Schlee, Kerry Smith and Galina Vereshchagina.

For many helpful conversations, I thank my classmates and friends, particularly Adam Blandin, Nastassia Krukava, Rishabh Sinha, Xin Tang and Jianhuan Xu. The second chapter of this dissertation is coauthored with Ji Qi at Tsinghua University and Xin Tang at Wuhan University.

Finally, I am deeply indebted to my friend and former advisor Deming Luo, who introduced me to economics and inspired me to pursue my passion in economics.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Chapter 1

MULTI-ESTABLISHMENT FIRMS, MISALLOCATION AND PRODUCTIVITY

1.1 Introduction

Misallocation of resources among production units can potentially cause large losses in aggregate productivity, and therefore can play a large role in explaining the income disparity across countries. [1] An important task then is to identify institutional and policy distortions that are more damaging to aggregate productivity, and therefore are more relevant in explaining the income disparity. A growing literature argues that size-dependent distortions, or distortions that are correlated with the size of production units, are a good candidate, because they reallocate resources from high-productivity to low-productivity production units. [2] Consistent with this story, there are large differences in the size of production units across countries. [3]

In this paper, I study the aggregate consequences of misallocation in a firm dynamics model with multi-establishment firms. In particular, I use the model to reassess the aggregate effects of size-dependent distortions to establishments, which are the focus of the previous studies. In contrast to previous findings, I find that size-dependent distortions are *not* more damaging to aggregate productivity than size-*in*dependent distortions, when they impose the same total burden on the whole business sector. The presence of multi-establishment firms plays a key role in driving this result.

While the typical analysis in the literature assumes that firms operate only one es-

[1] Hsieh and Klenow (2009), Restuccia and Rogerson (2008), and Guner *et al.* (2008) among others. See Restuccia and Rogerson (2013) for a review of the recent literature.

[2] Restuccia and Rogerson (2008) and Guner *et al.* (2008) among others.

[3] Bento and Restuccia (2014) and Lagakos (2016) document large cross-country differences in the size of manufacturing and retail establishments respectively.

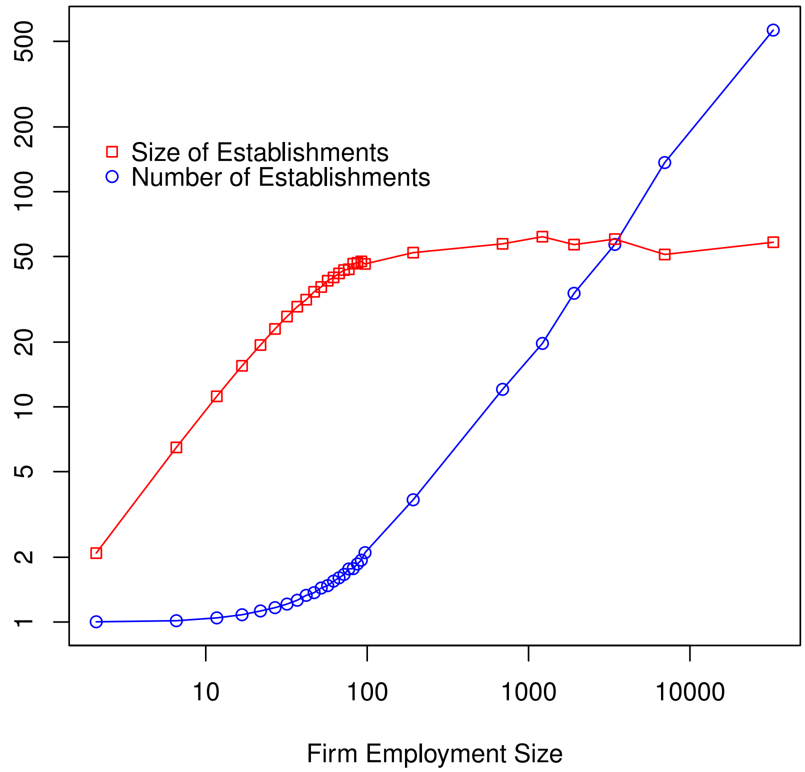

Figure 1.1: The Number and Employment Size of Establishments in the U.S. Firms of Different Employment Sizes

tablishment, most large firms in advanced economies operate multiple establishments. Figure 1 shows the average number and average employment size of establishments in the U.S. of firms with different employment sizes in 2003. Among firms with at least 100 employees, the average number of establishments rises almost log-linearly with firm employment size, while the average employment size of establishments hardly changes with firm employment size. This suggests that the extensive margin of creating additional establishments is crucial for the growth of large firms.

The presence of the extensive margin of creating additional establishments has important implications for the assessment of the aggregate consequences of misallocation. To begin with, changing the number of establishments provides large firms with an additional margin of adjustment when they face distortions. This margin is important if there are size-dependent distortions to establishments, for example, policies that limit the size of manufacturing establishments in India and the size of retail establishments in Japan as in Guner *et al.* (2008), because firms can react to them by operating a larger number of establishments with smaller size.

Moreover, the presence of the extensive margin is crucial for understanding the forces that drive the growth of large firms. When large firms grow by creating new establishments, the growth is not only driven by their investment in new intangible capital, such as the creation of new patents and new blueprints, but also by utilization of their existing intangible capital in new locations. Without taking the extensive margin of firm growth into account, we may overstate large firms' investment in intangible capital. The correlation between firm size and investment in new intangible capital is important for the effects of size-dependent distortions on aggregate productivity. If large firms invest more in intangible capital, size-dependent distortions not only cause static misallocation of resources given the productivity distribution, but also discourage investment in intangible capital. However, if small firms invest more in intangible capital, then by restricting the operations of large establishments and firms, size-dependent distortions make room for more small firms with high investment in intangible capital, which would largely offset the negative effects on aggregate productivity from the static misallocation of resources.

In addition, various institutional and policy distortions induce misallocation of resources among firms instead of establishments. For example, financial frictions are likely to affect the operations of the whole firm and labor unions are more likely to target establishments that are part of multi-establishment firms. [4] There are also

[4]See Dinlersoz *et al.* (2014).

restrictions on the creation of establishments by multi-establishment firms. Notable examples include geographic restrictions on the US banking industry before the 1990s, which limited the ability to choose branch locations, and restrictions on the entry of large international retail chains in India's retail industry. These restrictions can potentially cause substantial losses to aggregate productivity since they tend to restrict the operations of productive firms, but standard heterogeneous firm models are not well suited to studying their consequences. [5]

Motivated by the potential importance of multi-establishment firms for misallocation and aggregate productivity, I build a model of multi-establishment firms based on the standard equilibrium model of firm dynamics along the lines of Hopenhayn (1992), Hopenhayn and Rogerson (1993) and Veracierto (2001). I add two new elements to the standard model: technology capital and multi-establishment firms. Technology capital measures the stock of firm's unique know-how from investing in R&D, brands, and organization capital, which can be used simultaneously by all establishments operated by a firm. [6] Because it is non-rival, technology capital provides a rationale for the existence of multi-establishment firms.

My model has the following features. The establishment is the basic production unit in the model while the firm is the decision-making unit. In addition to labor and physical capital, establishments use two different types of intangible capital as inputs of production: technology capital that is non-rival and establishment-specific intangible capital that is rival. Firm also can choose to operate multiple establishments

[5] One motivation comes from the importance of resource reallocation from single-establishment to multi-establishment firms in aggregate productivity growth. For example, Foster *et al.* (2006) find that the entry of high-productivity establishments that belong to large national retail chains and the exit of low-productivity single-establishment firms have played a dominant role in aggregate productivity gains in the US retail trade sector.

[6] See McGrattan and Prescott (2009, 2010). Markusen (1984) develops a similar concept named knowledge capital, which measures the stock of firm's intangible assets that have a joint-input characteristic

at a cost. Hence, in addition to standard decisions, firms make two extra decisions: the amount of investment in intangible capital and the number of establishments in operation.

The non-rivalry of technology capital leads to the economies of multi-establishment operations: given the same amount of inputs, firms with more establishments produce more output because the technology capital they own can be used simultaneously by more establishments. This has important implications for firm and establishment sizes, and can be used to infer the importance of technology capital in firm production. If technology capital is more important in firm production, large firms will expand more along the extensive margin and operate more establishments, but the size of their establishments will be smaller.

I calibrate the model to match salient features of the firm size distribution with respect to both the number of employees and the number of establishments in the US. Despite its simplicity, the model does a surprisingly good job in matching the whole firm size distribution. As a by-product of the calibration, I obtain an estimate the importance of technology capital in production. It is reassuring that my estimate is roughly consistent with the findings in previous studies. [7] I use the calibrated model to evaluate the effects of size-dependent distortions to establishments, which are the focus of previous studies and modeled as implicit output taxes.

In contrast to the findings in previous studies, I find that size-dependent distortions are *not* more damaging to aggregate productivity and output than size-*in*dependent distortions, if they impose the same burden on the whole business sector, measured by the implicit tax revenue. In addition, the effects of distortions on aggregate output are approximately summarized by the implicit tax revenue from the distortions: distortions with different degree of size dependency have similar effects

[7] McGrattan and Prescott (2010) and Kapika (2012)

on aggregate output if the same amount of implicit tax revenue is collected. This result suggests an important implication for empirical studies on misallocation: in order to identify distortions that are more damaging to aggregate productivity and output, we should pay more attention to distortion that impose larger burden on the whole business sector, instead of distortions that are more correlated establishment size.

The presence of the extensive margin of creating additional establishments plays a key role in this result. It provides large firms with an additional margin of adjustment, and is crucial for model predictions on large firms' investment in intangible capital. To highlight the importance of the extensive margin for the result, I redo the above exercise in a model which is identical to my model except that there is no technology capital, so every firm operates only one establishment, and is calibrated to match the US firm size distribution with respect to number of employees only. I find that large firm invest more in intangible capital in that model, and size-dependent distortions are more damaging to aggregate productivity and output than size-*in*dependent.

I also use the model to compare effects of size-dependent distortions to establishments and to firms. I find that these two forms of distortions have very different effects on firm size distribution. Size-dependent distortions to establishments induce multi-establishment firms to have a larger number of smaller establishments, while size-dependent distortions to firms induce a much larger reduction in mean firm size and the number of establishments per firm. However, their effects on aggregate output are similar. This suggests another implication for empirical studies on misallocation: cross-country differences in firm size distributions alone do not tell us much about the distortionary effects on aggregate productivity and output.

In addition, I use the model to study restrictions on establishment creation, and find that they induce induce a large reduction in mean firm size and the number of establishments per firm. However, the effects of these restrictions on output depends on whether they apply to a small sector, in which case wage is approximately fixed, or apply to the whole economy in which case wage adjusts endogenously. Restrictions that apply to a smaller sector would cause a big drop in the output of that sector, while restrictions that apply to the whole economy have a small impact on the aggregate output.

The remainder of the paper proceeds as follows. Section 2 illustrates key features of technology capital and multi-establishment operations in a simple static model. Section 3 describes the formal model with firm dynamics. Section 4 discusses the calibration of the dynamic model. Section 5 analyzes size-dependent distortions to establishments and firms, and restrictions on establishment creation. Section 6 concludes. All figures and tables are at the end of the paper.

1.2 A Static Model

In this section, I build a static model of heterogeneous firms that captures the key features of technology capital and multi-establishment firms, and use the model to show explicitly how distortions affect firms' decisions, firm size distribution and aggregate productivity. In Section 3, I will generalize this static model to the full blown dynamic model that I will connect with the data.

1.2.1 Model Environment

There is a continuum of firms with measure normalized to 1. Each firm is endowed with a stock of intangible capital x, which is drawn from a probability density function $\phi(x)$. Firms use intangible capital x to produce technology capital z_f, which is non-rival within a firm and can be used simultaneously by all the establishments in a firm, and establishment-specific intangible capital z_e that is rival, using the following technology

$$z_f + z_e \leq x \tag{1.1}$$

Examples of technology capital z_f are blueprints and patents owned by a firm, and a firm's brands and reputations. Examples of establishment specific intangible capital z_e are local customer base of a specific establishment in a firm, and a manager's knowledge about local production conditions.

The establishment is the basic production unit in this economy, and firms can operate multiple establishments. An establishment uses technology capital z_f, establishment specific intangible capital z_e, and labor h to produce the final good y, according to the following production technology:

$$y = f(z_f, z_e, h) = (z_f^{\alpha} z_e^{1-\alpha})^{1-\gamma} h^{\gamma} \tag{1.2}$$

where $\alpha, \gamma \in (0, 1)$. α determines the importance of technology capital z_f in establishment production, and $1 - \gamma$ measures the importance of intangible capital. A firm needs to hire τ units of labor to create and operate an establishment.

Finally, the total endowment of labor in the economy is H, which is supplied inelastically.

1.2.2 Firm Optimization Problem

The firm optimization problem can be decomposed into two stages. In the second stage, firms solve an output maximization problem: given the amount of intangible capital x, the number of establishments n, and the total amount of labor h in a firm, the firm chooses $\{z_f, z_{e,j}, h_j\}_{j=1}^{n}$ to maximize the output it produces, where z_f, $z_{e,j}$ and h_j are technology capital, establishment specific intangible capital and labor used by establishment j of the firm, respectively. In Stage 1, the firm chooses the number of establishments n and the amount of labor h to maximize its profit.

In Stage 2, given the amount of production inputs $\{x, h\}$ and the number of establishments n, the firm chooses $\{z_f, z_{e,j}, h_j\}_{j=1}^{n}$ to solve the following output-maximization problem

$$F(x,n,h) = \max_{\{h_j,k_j\}} \sum_{j=1}^{n} f(z_f, z_{e,j}, h_j) \tag{1.3}$$

$$s.t. \sum_{j=1}^{n} h_j \leq h, \ \sum_{j=1}^{n} z_{e,j} \leq z_e, \ z_e + z_f \leq x$$

Notice that given the non-rival nature of technology capital, the firm does not need to allocate z_f among the establishments it operates. Given that all establishments have the same production function, the solution to this problem is

$$h_j = \frac{h}{n}, \ z_{e,j} = \frac{z_e}{n}$$

$$z_e = (1-\alpha)x, \ z_f = \alpha x$$

That is, all establishments in the firm use the same amount of establishment specific intangible capital and labor, and a share α of the intangible capital x is used to produce technology capital. Substituting the solution into (1.3), the production function of the firm can be written as

$$F(n,x,h) = \Omega(n^{\alpha}x)^{1-\gamma}h^{\gamma} \tag{1.4}$$

Unlike standard models of firm heterogeneity, the firm's output in this model is not only determined by the amount of intangible capital x, but also by the number of establishments it operates, n, which captures the scale economies of multi-establishment operations. Notice that the parameter which determines the importance of technology capital in establishment production, α, is precisely the parameter which determines the scale economies of multi-establishment operations. An immediate implication of this observation is that, data on firms' multi-establishment operations provide important information on the magnitude of α.

In Stage 1, given wage w, a firm with intangible capital x chooses $\{n, h\}$ to maximize its profit

$$\pi(x) = \max_{n,h} \Omega(n^\alpha x)^{1-\gamma} h^\gamma - wh - \tau wn$$

and the optimal solution is

$$h(x) = A(\alpha, \gamma) x^{\frac{1}{1-\alpha}} \tau^{\frac{\alpha}{\alpha-1}} w^{\frac{1}{(\alpha-1)(1-\gamma)}} \tag{1.5}$$

$$n(x) = B(\alpha, \gamma) x^{\frac{1}{1-\alpha}} \tau^{\frac{1}{\alpha-1}} w^{\frac{1}{(\alpha-1)(1-\gamma)}} \tag{1.6}$$

$$\frac{h(x)}{n(x)} = \frac{\gamma}{(1-\gamma)} \frac{\tau}{\alpha} \tag{1.7}$$

Quite intuitively, these equations say that both the amount of production labor $h(x)$ and the number of establishments $n(x)$ in the firm are increasing in x and decreasing in τ and w. That is, firms with a larger amount of intangible capital use more labor and operate more establishments. Moreover, the amount of labor used by each establishment, $h(x)/n(x)$, is independent of x, implying that differences in firm employment size are caused solely by differences in number of establishments: larger firms have more establishments, but not larger establishments. This seems at odds with the US data at first glance. However, if we focus on multi-establishment firms, the model provides a reasonable approximation of the data. Figure 1 shows the average number and employment size of establishments in firms of different employment sizes on a log-log plot. For firms with more than 100 employees, average establishment size does not rise or fall with firm employment size, which is consistent with what the model predicts.

1.2.3 Competitive Equilibrium

A competitive equilibrium in this economy consists of a list of $\{h(x), n(x)\}$ and a wage w such that

1. Given wage w, $\{h(x), n(x)\}$ solves the profit-maximization problem of a firm with intangible capital x

2. The labor market clears

$$\int (h(x) + \tau n(x))\phi(dx) = H \tag{1.8}$$

Notice that $\tau n(x)$ is included in aggregate labor demand because the costs of operating establishments are paid in terms of labor.

1.2.4 Size-Dependent Distortions

In this section, I use the model to evaluate the aggregate effects of size-dependent distortions to establishments and to firms, i.e., distortions correlated with establishment and firm sizes, respectively.

Size-dependent distortions are prevalent in both developed and developing countries. Guner *et al.* (2008) provide interesting examples of size-dependent distortions in developed countries, such as the regulation of the retail sector in Japan and the labor regulation in Italy. Using representative firm-level data in India, Indonesia and Mexico, Hsieh and Olken (2014) find large firms have higher average products of capital and labor, suggesting large firms in those countries face higher input prices. This particular type of size-dependent distortions, which restrict the operations of large production units, are also emphasized in Restuccia and Rogerson (2008) and Hsieh and Klenow (2014)). They tend to move resources from more to less productive units, therefore causing substantial losses in aggregate productivity.

In a heterogeneous firm model with only single-establishment firms, size-dependent distortions to establishments would produce the same aggregate effects as size-dependent distortions to firms. This is no longer the case in a model with multi-establishment firms. In the presence of distortions that restrict the operations of large establishments, firms that own multiple large establishments could respond by operating more establishments of smaller sizes, and the amount of resources used by the whole firm may not drop as much. In other words, size-dependent distortions to establishments cause more misallocation *within* firms, while size-dependent distortions to firms cause more misallocation *across* firms. My simple model can be used to illustrate these points.

Size-Dependent Distortions to Establishments

I study the effects of size-dependent distortions to establishments, which are the focus of previous studies such as Guner *et al.* (2008) and Restuccia and Rogerson (2008). I introduce the following taxes whose rates rise with establishment output to the benchmark economy: an establishment with output y faces a tax rate $1 - \kappa y^{-\rho}$, where $0 \leq \rho \leq 1$. This tax function was first proposed by Benabou (2002) and has been popular in the fields of development economics and public economics largely due to its simplicity: the parameter κ determines the mean level of the taxes, while ρ determines the size dependency of the taxes. When $\rho = 0$, all the establishments, small or large, face the same tax rate $1 - \kappa$. [8] In this case the tax scheme is size-independent. When $\rho > 0$, establishments with a larger amount of output y face higher tax rates.

The firm's optimal choice involves dividing the rival inputs evenly across its establishments. Therefore firm's problem can be written as

$$\pi(x) = \max_{n,h,z_f,z_e} \kappa n[\left(z_f^{\alpha} \left(\frac{z_e}{n}\right)^{1-\alpha} \right)^{1-\gamma} \left(\frac{h}{n}\right)^{\gamma}]^{1-\rho} - wh - w\tau n$$

where $z_f + z_e \leq x$. The solution is for any x and x',

$$\frac{h(x)}{h(x')} = \frac{n(x)}{n(x')} = \left(\frac{x}{x'}\right)^{\frac{1}{1-\alpha}} \tag{1.9}$$

[8] See Guner *et al.* (2015) for a discussion of recent studies that use this tax function.

$$\frac{h(x)}{n(x)} = \frac{\gamma(1-\rho)}{1-(\gamma+(1-\alpha)(1-\gamma))(1-\rho)} \tag{1.10}$$

Equation (1.9) says the ratio of labor used in two firms is independent of ρ and κ. Since aggregate labor supply is fixed, each firm would use the same amount of labor as in the undistorted economy, and there's no misallocation of labor across firms. However, as made clear by Equation (1.10), the establishment size $h(x)/n(x)$ is decreasing in ρ, so size-dependent distortions to establishments induce firms to expand more along the extensive margin and operate a larger number of smaller establishments. Therefore, in this simple model, size-dependent distortions to establishments cause misallocation *within* firms rather than *across* firms.

Size-Dependent Distortions to Firms

Now I introduce the following size-dependent distortions to firms to the benchmark economy. Specifically, a firm with output y faces a tax rate $1-\kappa y^{-\rho}$, where $0 \leq \rho \leq 1$. The parameter ρ determines the size dependency of the taxes: when $\rho = 0$, all the firms face the same tax rate $1-\kappa$, and when $\rho > 0$, firms with larger output y face higher tax rates.

Given the taxes and wage w, a firm with intangible capital x chooses $\{n, h\}$ to maximize its after-tax profit

$$\pi(x) = \max_{n,h} \kappa(\Omega(n^{\alpha}x)^{1-\gamma}h^{\gamma})^{1-\rho} - wh - w\tau n$$

The optimal solution is

$$h(x) = A \cdot x^{\frac{(1-\gamma)(1-\rho)}{\beta}} \tag{1.11}$$

$$\frac{h(x)}{n(x)} = \frac{\gamma\tau}{\alpha(1-\gamma)} \tag{1.12}$$

where $\beta = 1-(\gamma+\alpha(1-\gamma))(1-\rho)$.

As above, firms with a larger x use more labor and operate a larger number of establishments, as both $h(x)$ and $n(x)$ are increasing in x. In addition, all establishments in the economy use the same amount of production labor, as $h(x)/n(x)$ is independent of x, which is not affected by ρ as well.

However, for any x and x',

$$\frac{h(x)}{h(x')} = \left(\frac{x}{x'}\right)^{\frac{(1-\gamma)(1-\rho)}{1-(\gamma+\alpha(1-\gamma))(1-\rho)}} \tag{1.13}$$

$\frac{(1-\gamma)(1-\rho)}{1-(\gamma+\alpha(1-\gamma))(1-\rho)}$ is decreasing in ρ, implying that when ρ is larger, the ratio of the labor used by firms x and x' is smaller. Therefore, given the distribution of x, size-dependent distortions to firms cause a more compressed firm employment size distribution. To take stock, in contrast to size-dependent distortions to establishments, size-dependent distortions to firms cause misallocation *across* firms rather than *within* firms in this simple model. In the full blown dynamic model below, size-dependent distortions cause misallocation both *across* and *within* firms, but the quantitative results confirm the basic insight in this simple model: size-dependent distortions to establishments cause more misallocation *within* firms, while size-dependent distortions to firms cause more misallocation *across* firms.

To further exploit the effects of firm-level size-dependent distortions on aggregate output, I assume x is distributed according to a log-normal distribution with mean μ and standard deviation σ, i.e., $\phi(x)$ is a log-normal probability distribution function. One can show that the aggregate output in the economy then is given by

$$log(Y) = \Gamma(\alpha,\gamma,\tau) + (\gamma + (1-\gamma)\alpha)log(H) + (1-\gamma)\mu + \Psi(\rho,\alpha,\gamma)\frac{\sigma^2}{2} \tag{1.14}$$

where the function $\Psi(\rho,\alpha,\gamma)$ is strictly decreasing in ρ. Equation (1.14) has an intuitive interpretation: aggregate output in this economy is determined by the total amount of labor available in this economy, whose relative importance is $\gamma+(1-\gamma)\alpha$, by the mean value of the distribution of intangible capital, whose relative importance is $(1-\gamma)\mu$, and by the dispersion of firm size distribution, whose relative importance is a function of the importance of technology capital α, the size-dependency of distortions ρ and the dispersion of intangible capital σ. Since $\Psi(\rho, \alpha, \gamma)$ is strictly decreasing in ρ, size dependent distortions reduce aggregate output precisely by compressing the distribution of firm sizes.

Restrictions on Establishment Creation

The cost of operating an additional establishment, τ, is a key parameter of the model. In the real world, τ is affected by both technological and institutional factors, such as the communication cost between headquarters and affiliated establishments, the cost headquarter managers incur to monitor the managers of affiliated establishments, and the regulation of entry of establishments. [9] Technological advances in communication would reduce τ, while increases in the regulation of establishment entry would raise τ.

Due to poor communication technology and infrastructure and high regulation of entry, firms in developing countries usually face higher costs when they operate additional establishments. Another important factor that limits firms' multi-establishment operations is weak rule of law in developing countries. Bloom *et al.* (2013) find the number of male family members is the dominant factor determining the number of plants for Indian textile firms, which can be explained by weak rule of law in India. In addition, when doing business in regions far away from their headquarters, firms in developing countries may also have to pay large bribes to local government officials.

There are also government policies which increase the costs of establishment cre-

[9] See Grossman and Rossi-Hansberg (2008) and Head and Ries (2008)

ation. One example comes from the banking industry in the US. Historically, banks in the US were limited in their ability to choose branch office locations. As late as 1974, only 14 states allowed statewide branching and 12 states prohibited branching altogether, while the rest of the states allowed limited branching only. [10] The banks faced even more severe restrictions when they wanted to own and operate branches across state lines. The McFadden Act of 1927 specifically prohibited banks from interstate branching, until the provision was repealed by the Riegle-Neal Interstate Banking and Branching Efficiency Act of 1994. Another example comes from India's retail industry, a rapidly growing industry that accounts for more than 20% of India's GDP. Retail industry in India is characterized by small family-owned stores with low labor productivity, which account for more than 90% of the market share. Meanwhile, entry by large international retail chains is restricted, especially in the multi-brand retail category. Until 2011, FDI was denied in multi-brand retail, and large international retail chains like Wal-Mart and Carrefour were not allowed to operate supermarkets or retail outlets in India. The model in this paper is well-suited to studying the effects of these distortions.

I model the restrictions on establishment creation as an increase in τ, and study the effects of restrictions that apply to a small sector, in which case wage w is determined exogenously and will not be affected by the restrictions, and that restrictions that apply to the whole economy, in which case wage w is determined endogenously and will be affected by the restrictions.

I start with restrictions that apply to a small sector. Since the wage w is not affected when the sector is small, according to Equations (1.5) and (1.6), both the amount of production labor $h(x)$ and the number of establishments $n(x)$ drop for each firm when τ increases. Meanwhile, from Equation (1.7) we see that the mean

[10] Jayaratne and Strahan (1997)

employment size of establishments $h(x)/n(x)$ rises. The intuition is, as the cost of multi-establishment operations increases, firms choose to operate a smaller number of establishments with larger employment sizes. As a result, total output Y in the sector drops:

$$\frac{\partial ln(Y)}{\partial ln(\tau)} = -\frac{\alpha}{1-\alpha}$$

which implies that output drops more when α is larger, i.e., when technology capital is more important for firm production.

Now I study restrictions that apply to the whole economy, in which case wage w is determined endogenously by the labor market clearing condition and will be affected by the restrictions. As before, the number of establishments $n(x)$ drops while the mean employment size of establishments $h(x)/n(x)$ rises for each firm when τ increases. However, given the fixed measure of firms and fixed aggregate labor supply, the wage w will adjust in such a way that each firm will use the same amount of labor as in the benchmark economy. Since firms operate an inefficiently smaller number of establishments, aggregate output Y now drops by

$$\frac{\partial ln(Y)}{\partial ln(\tau)} = -\alpha(1-\gamma)$$

Again, output drops more when α is larger. However, the output drops by a smaller amount than the previous case for the same value of τ, which reflects the importance of the general equilibrium effects.

Notice that in this simple model, these conclusions do not depend on the distribution of intangible capital x.

1.3 A Dynamic Model

In this section, I make previous model dynamic and introduce two important new elements to it: firm entry and exit; firm growth caused by intangible capital accumulation. Later I will use the dynamic model for quantitative analysis.

1.3.1 Establishments and Firms

Establishment is the basic production unit in this economy, which uses technology capital z_f, establishment specific intangible capital z_e, and physical capital k and labor h to produce the final good y, according to the following production technology:

$$y = f(z_f, z_e, k, h) = (z_f^{\alpha} z_e^{1-\alpha})^{1-\gamma}(k^{\eta} h^{1-\eta})^{\gamma}$$

where α determines the importance of technology capital z_f in establishment production, $1-\gamma$ measures the importance of intangible capital, and η determines the relative importance of physical capital to labor.

A firm in each period is characterized by the amount of intangible capital it owns, x. Intangible capital x can be used to produce technology capital z_f and establishment- specific intangible capital z_e, with the following technology

$$z_f + z_e \leq x$$

Firms can accumulate intangible capital x over time. If a firm with x uses I_t units of labor for the accumulation of intangible capital, then in the next period the stock of intangible capital would be

$$x_{t+1} = [(1-\delta_x)x_t + x_t^{\theta_1} I_t^{\theta_2}] \exp(\epsilon_{t+1}) \tag{1.15}$$

δ_x is the depreciation rate of intangible capital and ϵ_{t+1} is a shock whose value is realized at the beginning of period $t+1$. As is evident from (1.15), both the current stock of intangible capital x_t and the investment on intangible capital I_t play a role in the production of intangible capital, where θ_1 determines the importance of the current stock of intangible capital in the production of new intangible capital, while θ_2 determines the importance of investment I_t.

Firms can choose to operate multiple establishments. Firms need to use τ_e units of labor to operate an additional establishment and this cost is not paid if firms operate only one establishment. Therefore, the total operating cost to a firm n establishments is

$$oc(n) = \begin{cases} (n-1)\tau_e & n > 1 \\ 0 & n = 1 \end{cases}$$

The parameter τ_e captures the regulation of establishment entry, the communication cost between headquarters and affiliate establishments and the cost of monitoring establishment managers and so forth.

There is an unlimited mass of potential entering firms. In each period, a new firm can enter the economy after paying a sunk cost of τ_f measured in terms of labor. Upon entry, the firm draws its initial level of intangible capital from a distribution $\phi(x)$. The draws are i.i.d. across firms. Finally, with probability λ each firm is hit by a shock that forces it to exit.

1.3.2 Household

There is a stand-in household that consists of a continuum of members with measure normalized 1. The household is infinitely lived and maximizes the lifetime utility

$$\sum_{t=0}^{\infty} \beta^t U(C_t)$$

where C_t is consumption of the final good in period t.

In each period, the household is endowed with one unit of productive time, which it supplies inelastically. The household owns the physical capital and firms.

1.3.3 Stationary Equilibrium

I focus on a stationary equilibrium in which the equilibrium wage w_t, the rental price for physical capital R_t and interest rate r_t are constant over time.

Incumbent Firm's Problem

As in the static model, a firm's static problem can be broken down into two stages. In the second stage, the firm solves an output maximization problem: given intangible capital x, the number of establishments n, and the amount of labor h and physical capital k, the firm chooses inputs $\{z_f, z_{e,j}, k_j, h_j\}_{j=1}^{n}$ to maximize its output, where $(z_f, z_{e,j}, k_j, h_j)$ are the inputs used by establishment j in the firm. In Stage 2, the firm chooses the number of establishments n , the amount of labor h and physical capital k to maximize its profit.

In Stage 2, given the amount of production inputs $\{x, k, h\}$ and the number of establishments n in the firm, the firm chooses $\{z_f, z_{e,j}, k_j, h_j\}_{j=1}^{n}$ to solve the following problem

$$F(x, n, k, h) = \max_{\{h_j, k_j\}} \sum_{j=1}^{n} f(z_f, z_{e,j}, k_j, h_j)$$

$$s.t. \sum_{j=1}^{n} k_j \leq k, \sum_{j=1}^{n} h_j \leq h$$

$$\sum_{j=1}^{n} z_{e,j} \leq z_e, \; z_e + z_f \leq x$$

Notice that given the non-rival nature of technology capital, the firm does not need to allocate z_f among the establishments it operates. Given that all establishments have the same production function, the solution to this problem is

$$k_j = \frac{k}{n}, \; h_j = \frac{h}{n}, \; z_{e,j} = \frac{z_e}{n}$$

$$z_e = (1-\alpha)x, \; z_f = \alpha x$$

Hence the firm production function is

$$F(x,n,k,h) = \Omega(n^{\alpha}x)^{1-\gamma}(k^{\eta}h^{1-\eta})^{\gamma}$$

As before, the firm's output is determined not only by the amount of intangible capital x, but also by the number of establishments it operates, which captures the scale economies of multi-establishment operations, and α determines the scale economies of multi-establishment operations.

In Stage 1, given wage w and rental rate of capital R, a firm with intangible capital x chooses $\{n,k,h\}$ to maximize its profit

$$\pi(x) = \max_{n\geq 1,k,h}\{\Omega(n^{\alpha}x)^{1-\gamma}(k^{\eta}h^{1-\eta})^{\gamma} - wh - Rk - w\tau_e(n-1)\}$$

There is a cutoff level $\hat{x}$ such that firms with $x < \hat{x}$ choose to operate only one establishment, for any $x, x' < \bar{x}$

$$\frac{h(x)}{h(x')} = \left(\frac{x}{x'}\right) \tag{1.16}$$

while firms with $x \geq \bar{x}$ choose to operate multiple establishments, and any $x, x' > \bar{x}$

$$\frac{h(x)}{h(x')} = \left(\frac{x}{x'}\right)^{\frac{1}{1-\alpha}} \tag{1.17}$$

In addition, mean establishment sizes in multi-establishment firms are

$$\frac{h(x)}{n(x)} = \frac{\gamma(1-\eta)}{(1-\gamma)}\frac{\tau_e}{\alpha} \tag{1.18}$$

Figure 2 shows the model predictions on mean establishment size and number of establishments in firms of different employment sizes. Comparing Figure 2 with Figure 1, which depicts mean establishment size and number of establishments in firms of different employment sizes in the US data, we find that the model predictions capture the key features of the data: most small firms are single-establishment firms, while

Figure 1.2: The Number and Employment Size of Establishments in Firms of Different Employment Sizes in the Model: A Numerical Example

large firms typically operate multiple establishments, and average employment size of establishment is roughly constant for those large firms.

Let $V(x)$ be the value of a firm with intangible capital x, then

$$V(x) = \max_{n,h,k,I} \{F(x,n,h,k) - wh - Rk - w\tau_e(n-1) - wI + \frac{1-\lambda}{1+r} E[V(x')|I,x]\} \tag{1.19}$$

where

$$x' = [(1-\delta_x)x + x^{\theta_1} I^{\theta_2}] \exp(\epsilon') \tag{1.20}$$

Entering Firm's Problem

Entering firms pay a sunk cost of τ_f units of labor, and then draw an initial level of intangible capital from distribution $\phi(x)$. Let V_E be the present discounted value of a potential entering firm:

$$V_E = \int V(x)\phi(dx) - w\tau_f \tag{1.21}$$

where τ_f denotes the entry cost.

In an equilibrium with firm entry, V_E must be equal to zero since otherwise additional firms would enter. Following the literature, I refer to the condition $V_E = 0$ as the free-entry condition. Since V_E is strictly decreasing in wage w, it follows that there is a unique value of w for the free-entry condition to hold. Therefore, if there is an equilibrium with firm entry, then the free-entry condition will determine the wage rate w.

Household's Problem

Given the wage w_t and the rental price of physical capital R_t, the household's problem can be written as

$$\max_{\{C_t, K_t\}} \sum_{t=0}^{\infty} \beta^t U(C_t)$$

subject to

$$C_t + K_{t+1} - (1-\delta)K_t \leq w_t + R_t K_t + \Pi_t$$

where K_t is the total stock of physical capital, which depreciates at rate δ, and Π_t is the aggregate profits of all the incumbent firms in the economy in period t.

In a stationary equilibrium, the Euler equation determines the rental price of physical capital R

$$R = \frac{1}{\beta} - 1 + \delta$$

It follows that in a stationary equilibrium, the interest rate r is determined solely by the discount factor β

$$r = R - \delta = \frac{1}{\beta} - 1$$

Definition of A Stationary Equilibrium

A stationary competitive equilibrium with firm entry consists of a wage rate w, a real interest rate r, a rental price of physical capital R, a distribution of firms $\mu(x)$, a mass of firm entry E, a value function for incumbent firms $V(x)$ and a value of entering firms V_E , policy functions $\{n(x), k(x), h(x), I(x)\}$ for incumbent firms, and aggregate levels of consumption C and physical capital K such that the following conditions are satisfied:

1. Consumer optimization:
$$r = \frac{1}{\beta} - 1$$
2. Firm optimization: Given prices (w, R) , the function $V(x)$ solves the incumbent firms' problem and $\{n(x), k(x), h(x), I(x)\}$ are optimal policy functions.
3. Free-entry condition:
$$V_E = \int V(x)\phi(dx) - w\tau_f \tag{1.22}$$
4. Market clearing:
$$\int [h(x) + n(x)\tau_e + h(x)]d\mu(x) + E\tau_f = 1$$
$$\int k(x)d\mu(x) = K$$
$$C + \delta K = \int F(x, n(x), k(x), h(x))d\mu(x)$$
5. μ is an invariant distribution.

1.4 Calibration

I assume the US is a distortion-free economy and calibrate the model parameters to match the key aspects of the US data at the aggregate and cross section levels.

Some parameters are chosen to match the aggregate level data in the US. A model period is chosen to be one year. Following Guner *et al.* (2008), I assume the stock of physical capital consists of business equipment and structures, business inventories and business land. For the period 1960–2000, the physical capital to measured GDP ratio averaged about 2.33, the share of physical capital in measured GDP is 0.32, while the ratio of investment in physical capital to GDP is 0.14. The discount factor β is chosen to match the physical capital to GDP ratio, and the depreciation rate of physical capital δ is chosen to match the ratio of investment in physical capital to GDP. The parameter that determines the relative importance of physical capital in production, η, joint with the span-of-control parameter γ, determines the share of physical capital in GDP. The literature has identified a narrow range of possible values for γ. [11] From that range I choose $\gamma = 0.8$ and given that choice, I choose η to match the share of physical capital in GDP.

The remaining parameters are chosen to match salient features of firm dynamics and size distribution in the US. First, the amount of job creation and destruction at the firm level is large. Second, large firms are relatively few in number, but account for a disproportionately large share of total employment. For example, establishments with 100 or more employees represent less than 3% of the total number of establishments but account for 46% of total employment, and firms with 500 or more employees represent only 0.4% of the total number of employer firms but account for almost 50% of total employment. Finally, as made clear by Figure 1, most small firms

[11] See the discussion in Guner *et al.* (2008).

are single-establishment firms, while large firms typically operate multiple establishments. Moreover, the average employment size of establishment is roughly constant for those large firms. The model in this paper is qualitatively consistent with these features, and model parameters are chosen to quantitatively match data moments concerning these features.

The stock of intangible capital is a key determinant of firm employment sizes in the model. As mentioned above, it evolves according to the following equation

$$x_{t+1} = [(1-\delta_x)x_t + x_t^{\theta_1} I_t^{\theta_2}] \exp(\epsilon_{t+1}) \tag{1.23}$$

where x_t is the current stock of intangible capital and I_t is the investment in intangible capital in period t . δ_x is the depreciation rate of intangible capital, θ_1 determines the importance of the current stock of intangible capital in the production of new intangible capital, while θ_2 determines the importance of investment I_t. ϵ_{t+1} is a random shock whose value is realized at the beginning of period $t+1$. I assume ϵ_{t+1} is distributed according to a normal distribution with mean 0 and standard deviation σ.

I choose (θ_1, θ_2) to match some important moments of the firm employment size distribution in the US, namely, mean firm employment size and employment share of firms with more than 500 employees. I use the 2003 data. The results don't change substantially if I use the data from other years in recent decades, as firm size distribution has been quite stable since 1970s in the US. As we see in Figure 3, the model does a good job in matching the entire firm size distribution. The standard deviation of shocks to intangible capital, σ, is chosen to match the average volatility of firm employment growth rates (employment weighted, excluding firm entry and exit) reported in Davis *et al.* (2007). The depreciation rate of intangible capital δ_x is

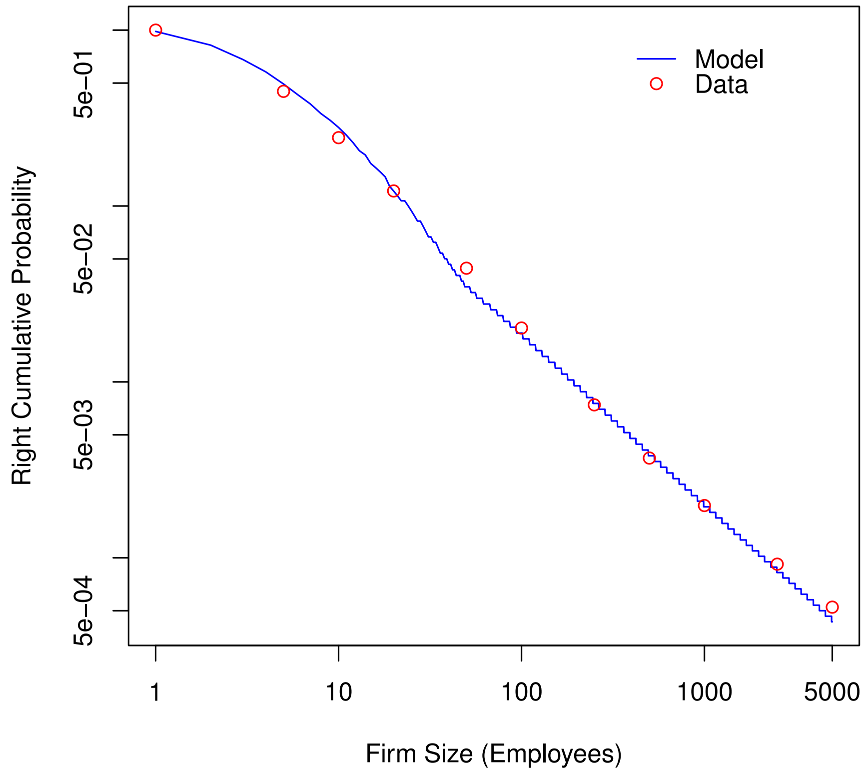

Figure 1.3: Firm Size Distribution: Model and Data

chosen to be 10% per year. [12]

As is made clear above, the parameter that determines the importance of technology capital in firm production, α, also determines the scale economies of multi-establishment operations. Other things being equal, when α is larger, firms with a larger stock of intangible capital x operate more but smaller establishments. In contrast, when the cost of operating additional establishments τ_e is larger, firms with a

[12]Empirical estimates of intangible capital depreciation rate are hard to come by. However, a large part of intangible capital is generated through investment in R&D, and empirical studies on R&D depreciation rate such as Schankerman and Pakes (1986) find the rate is above 10%. As a robustness check, I also try other values for depreciation rate of intangible capital, and the main results do not depend critically the choice of the depreciation rate.

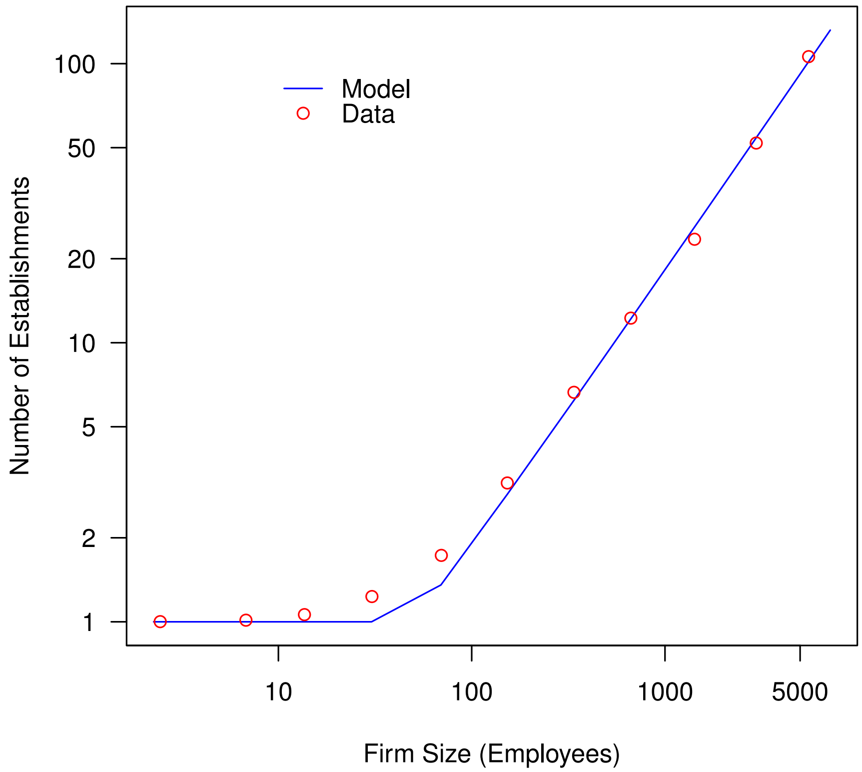

Figure 1.4: The Number of Establishments in Firms of Different Employment Sizes: Model and Data

larger stock of intangible capital x operate fewer but larger establishments. Therefore, I choose (α, τ_e) to match the average number of establishments owned by firms in the US economy, and average employment size of establishments in firms with 500 or more employees.

The distribution of initial draws of intangible capital, $\phi(x)$, is assumed to be a Pareto distribution with scale parameter x_m and shape parameter a. The reason behind this assumption is the size distribution of new firms is well approximated by a Pareto distribution. I choose the shape parameter a to match the shape parameter of the size distribution of new firms, and choose x_m to match the mean employment

	Parameters	Values
β	discount factor	0.93
η	importance of physical capital	0.40
δ	depreciation rate of physical capital	0.14
γ	span of control parameter	0.80
δ_x	depreciation rate of intangible capital	0.10
α	importance of technology capital in production	0.38
τ_e	cost of operating an establishment	7.13
θ_1	importance of current intangible capital	0.55
θ_2	importance of intangible capital investment	0.24
σ	volatility of shocks to intangible capital	0.22
a	shape parameter of distribution of initial x	1.25
x_m	scale parameter of distribution of initial x	2.72
λ	probability of firm death	0.08
τ_f	firm entry cost	13.10

† This table shows calibrated parameters for the benchmark economy.

Table 1.1: Calibration: Parameter Values

size of new firms.

Finally, the probability of firm death shock, λ, is chosen to match the mean US firm exit rate in 2003. The equilibrium wage in the benchmark economy is normalized to 1, and firm entry cost τ_f is chosen so that the free-entry condition (1.21) is satisfied. I summarize the parameter values and the targets in Tables (1) and (2).

Targets	Data	Model
Physical capital to output ratio	2.33	2.34
Share of physical capital in output	0.32	0.31
Investment in physical capital	0.14	0.14
Mean firm size	22.6	22.5
Mean number of establishments per firm	1.27	1.25
Mean establishment size in firms w/ 500+ workers	53.9	53.8
Employment share of firms w/ 500+ workers	0.50	0.50
Firm exit rate	0.08	0.08
Mean size of new firms	5.8	5.9
Shape of size distribution of new firms	1.25	1.25

† This table shows the model and data moments used in calibration.

Table 1.2: Calibration: Model and Data Moments

1.5 Distortions

1.5.1 Size-Dependent Distortions

In this subsection, I use the calibrated model to evaluate the effects of firm-level and establishment-level distortions. In the presence of multi-establishment firms, these two types of distortions may produce different cross-section and aggregate effects.

Size-Dependent Distortions to Establishments

First, I study the cross-sectional and the aggregate effects of size-dependent distortions to establishments, which are the focus of previous studies such as Guner *et al.* (2008) and Restuccia and Rogerson (2008). I introduce the following taxes whose rates rise with establishment output to the benchmark economy: an establishment with output y faces a tax rate $1 - \kappa y^{-\rho}$, where $0 \leq \rho \leq 1$. The parameter κ determines the mean level of the taxes, while ρ determines the size dependency of the taxes. When $\rho = 0$, all the establishments,small or large, face the same tax rate $1-\kappa$. In this case the tax scheme is size-independent. When $\rho > 0$, establishments with a larger amount of output y face higher tax rates.

The static profit-maximization problem of a firm with intangible capital x now becomes

$$\pi(x) = \max_{n\geq 1, h, z_f, z_e} \kappa n[\left(z_f^{\alpha} \left(\frac{z_e}{n}\right)^{1-\alpha}\right)^{1-\gamma} \left(\left(\frac{k}{n}\right)^{\eta} \left(\frac{h}{n}\right)^{1-\eta}\right)^{1-\gamma}]^{1-\rho} - wh - Rk - w\tau_e(n-1)\}$$

where $z_f + z_e \leq x$.

There is a cutoff level $\bar{x}$ such that firms intangible capital $x < \bar{x}$ choose to operate only one establishment, and for any $x, x' < \bar{x}$

$$\frac{h(x)}{h(x')} = \left(\frac{x}{x'}\right) \tag{1.24}$$

while firms with $x \geq \bar{x}$ choose to operate multiple establishments, and for any $x, x' > \bar{x}$

$$\frac{h(x)}{h(x')} = \left(\frac{x}{x'}\right)^{\frac{1}{1-\alpha}} \tag{1.25}$$

$$\frac{h(x)}{n(x)} = g(\alpha, \rho)\gamma\tau_e \tag{1.26}$$

where $g(\alpha, \rho)$ is decreasing in ρ.

Equation (1.26) implies that the size of establishments in multi-establishment firms is decreasing in ρ: since establishments in multi-establishment firms have larger output, after the introduction of taxes whose rates are higher for establishments with larger output, those firms respond by operating a larger number of smaller establishments. Therefore, size-dependent distortions at establishment level induce multi-establishment firms to operate establishments of inefficiently smaller sizes. On

Benchmark					
Level (κ)	1	1	1	1	1
Size Dependency (ρ)	0	0.02	0.04	0.06	0.08
Statistics					
Aggregate Output	100	97.2	94.5	91.9	89.3
Wage	100	90.6	82.7	76.2	70.6
Mean Firm Size	100	73.4	56.2	43.7	35.0
Mean Establishment Size	100	76.4	60.6	48.9	32.2
Total Number of Firms	100	136.2	177.9	228.9	285.3
Total Number of Establishments	100	130.9	165.0	204.6	310.5
Establishments Per Firm	100	96.0	92.5	89.6	87.0
Employment Share of Firms 500+	50.0%	42.0%	35.3%	29.1%	23.0%

† This table shows the effects on displayed variables associated to establishment-level size-dependent distortions across steady states. Column 2 reports the values of displayed variables in the benchmark economy, most of which are normalized to 100. Column 3-6 report the changes from increasing the size dependency of distortions.

Table 1.3: Effects of Establishment-Level Size-Dependent Distortions

the other hand, Equations (1.24) and (1.25) imply that for a *given* distribution of intangible capital x across firms, size-dependent distortions at establishment level may not create much misallocation of capital and labor across firms; but since distortions affect firm's investment on intangible capital, they also affect the distribution of intangible capital x across firms.

To evaluate the effects of establishment-level size-dependent distortions, I compute stationary equilibria for $\kappa = 1$ and different levels of ρ: $\rho = 0.02$, 0.04, 0.06 and 0.08, and summarize the results in Table (3). As we can see, size-dependent distortions at establishment level have a large impact on firm size distribution: mean firm size, mean number of establishments per firm and the employment share of firms with more than 500 workers all decrease with ρ, and when $\rho = 0.08$, mean firm size is only a third of that in the benchmark economy, while mean number of establishments per firm drops by 13% and the employment share of firms with more than 500 workers drops from 50% in the benchmark economy to 23%. On the other hand, the distortions cause substantial losses to aggregate output: aggregate output decreases with ρ and when $\rho = 0.08$, aggregate output drops by 11% compared with the benchmark case. Notice, however, total tax revenue is not the same across these experiments.

Size-Dependent Distortions to Firms

I evaluate the effects of size-dependent distortions to firms, by introducing the following size-dependent output taxes to the benchmark economy: a firm with output y faces a tax rate $1 - \kappa y^{-\rho}$, where $0 \leq \rho \leq 1$. The parameter ρ determines the size dependency of the taxes while κ determines the level of the taxes.

Then the static profit-maximization problem of a firm with intangible capital x becomes

$$\pi(x) = \max_{n \geq 1, k, h} \{\kappa \Omega [(n^{\alpha} x)^{1-\gamma} (k^{\eta} h^{1-\eta})^{\gamma}]^{\rho} - wh - Rk - w\tau_e (n-1)\}$$

There is a cutoff level $\tilde{x}$ such that firms with intangible capital $x < \tilde{x}$ choose to operate only one establishment, and for any $x, x' < \tilde{x}$

$$\frac{h(x)}{h(x')} = \left(\frac{x}{x'}\right)^{d_1(\rho)} \tag{1.27}$$

while firms with $x \geq \tilde{x}$ choose to operate multiple establishments, and any $x, x' > \tilde{x}$

$$\frac{h(x)}{h(x')} = \left(\frac{x}{x'}\right)^{d_2(\rho)} \tag{1.28}$$

	Benchmark				
Level (κ)	1	1	1	1	1
Size Dependency (ρ)	0	0.02	0.04	0.06	0.08
Statistics					
Aggregate Output	100	97.7	94.5	91.4	88.7
Wage	100	87.5	79.3	73.2	68.2
Mean Firm Size	100	55.7	40.5	32.7	27.5
Mean Establishment Size	100	65.3	49.8	40.8	34.3
Total Number of Firms	100	179.5	246.9	305.8	363.6
Total Number of Establishments	100	153.1	200.8	245.1	291.5
Establishments Per Firm	100	84.8	81.6	80.4	80.0
Employment Share of Firms 500+	50.0%	16.0%	4.2%	1.3%	0.4%

† This table shows the effects on displayed variables associated to firm-level size-dependent distortions across steady states. Column 2 reports the values of displayed variables in the benchmark economy, most of which are normalized to 100. Column 3-6 report the changes from increasing the size dependency of distortions.

Table 1.4: Effects of Firm-Level Size-Dependent Distortions

$$\frac{h(x)}{n(x)} = \frac{\gamma(1-\eta)}{(1-\gamma)}\frac{\tau_e}{\alpha} \tag{1.29}$$

both $d_1(\rho)$ and $d_2(\rho)$ are decreasing in ρ. From Equation (1.27) and (1.28), for a given distribution of intangible capital x, size-dependent distortions produce a more compressed firm size distribution, while from Equation (1.29), the size of establishments in multi-establishment firms is not affected by distortions. Hence, firm-level size-dependent distortions induce firms with high x to use less labor and operate fewer establishments, but not smaller establishments.

To facilitate the comparison with the establishment-level size-dependent distortions, I compute stationary equilibria for $\kappa = 1$ and $\rho = 0.02$, 0.04, 0.06 and 0.08, and summarize the results in Table (4). Size-dependent distortions at firm level have an even larger impact on the firm size distribution: for each value of ρ, the decrease in mean firm size, mean number of establishments per firm and the employment share of firms with more than 500 workers are larger than they were for the same distortions at establishment level. For example when $\rho = 0.08$, mean firm size is only a fourth (vs. a third with distortions at establishment level) of that in the benchmark economy, while mean number of establishments per firm drops by 20% (vs. 13%) and the employment share of firms with more than 500 workers drop from 50% in the benchmark economy to 0.4% (vs. 23%). Nonetheless, the effects on aggregate output are similar for these two forms of distortions. For example, when $\rho = 0.08$, aggregate output drops by 11.3% (vs. 10.7%) compared with the benchmark case. Again, total tax revenue is not the same across these experiments.

Size-Dependent Distortions under Revenue Neutrality

An important feature of the previous experiments is that the amount of resources extracted via the distortions is not constant across them. To evaluate the quantitative importance of total tax revenue, I redo the above experiments under the constraint of revenue neutrality.

I start with establishment-level distortions with $\rho = 0.02$, and choose κ such that the total amount of resources extracted is 10% of aggregate output in the new steady state. For all the other distortions, choose κ such that the amount of resources extracted is the same as in the above case. I summarize the effects of establishment-level size-dependent distortions in Table (5), and the effects of firm-level size-dependent distortions in Table (6).

Benchmark					
Level (κ)	1	0.97	1.033	1.09	1.14
Size Dependency (ρ)	0	0.02	0.04	0.06	0.08
Statistics					
Aggregate Output	100	95.9	95.9	95.6	94.9
Wage	100	86.6	86.7	86.2	85.0
Mean Firm Size	100	73.4	56.2	43.7	35.0
Mean Establishment Size	100	76.4	60.6	48.9	32.2
Total Number of Firms	100	136.2	177.9	228.9	285.3
Total Number of Establishments	100	130.9	165.0	204.6	310.5
Establishments Per Firm	100	96.0	92.5	89.6	87.0
Employment Share of Firms 500+	50.0%	42.0%	35.3%	29.1%	23.0%

† This table shows the effects on displayed variables associated to establishment-level size-dependent distortions across steady states. Column 2 reports the values of displayed variables in the benchmark economy, most of which are normalized to 100. Column 3-6 report the changes from increasing the size dependency of distortions. The amount of resources extracted via the distortions is the same in Column 3-6.

Table 1.5: Effects of Establishment-Level Distortions: Revenue Neutrality

As we can see Table (5) and (6), these distortions have vastly different effects on firm size distribution, but their effects on aggregate output do not differ much. Let's focus on establishment-level size-dependent distortions with $\rho = 0.02$ and $\rho = 0.08$. When $\rho = 0.02$, compared with the benchmark economy, mean firm size drops by 26%, mean number of establishments per firm drops by 4% and the employment share of firms with more than 500 workers drop to 42%; when $\rho = 0.08$, mean firm size drops by 65%, mean number of establishments per firm drops by 13% and the

	Benchmark				
Level (κ)	1	0.985	1.038	1.083	1.125
Size Dependency (ρ)	0	0.02	0.04	0.06	0.08
Statistics					
Aggregate Output	100	97.0	96.2	94.9	93.5
Wage	100	85.6	83.7	82.0	80.6
Mean Firm Size	100	55.7	40.5	32.7	27.5
Mean Establishment Size	100	65.3	49.8	40.8	34.3
Total Number of Firms	100	179.5	246.9	305.8	363.6
Total Number of Establishments	100	153.1	200.8	245.1	291.5
Establishments Per Firm	100	84.8	81.6	80.4	80.0
Employment Share of Firms 500+	50.0%	16.0%	4.2%	1.3%	0.4%

† This table shows the effects on displayed variables associated to firm-level size-dependent distortions across steady states. Column 2 reports the values of displayed variables in the benchmark economy, most of which are normalized to 100. Column 3-6 report the changes from increasing the size dependency of distortions. The amount of resources extracted via the distortions is the same in Column 3-6.

Table 1.6: Effects of Firm-Level Distortions: Revenue Neutrality

employment share of firms with more than 500 workers drop to 23%. However, their effects on aggregate output are rather similar: when $\rho = 0.02$, aggregate output drops by 5% relative to the benchmark case, and when $\rho = 0.08$, aggregate output drops by 6%.

I also carry out another set of experiments to illustrate the quantitative importance of the amount of resources extracted via the distortions. For each size-dependent distortion, I compute the effects of size-independent (purely proportional) distortion that extracts the same amount of resources. I summarize the effects of size-dependent and the corresponding size-independent distortions in Table (7). As we can see, size-dependent and the corresponding size-independent distortions have very different effects on firm size distribution: unlike size-dependent distortions, size-independent distortions have a very small impact on mean firm sizes, mean number of establishments per firm and the employment share of large firms. However, both types of distortions have similar effects on aggregate output when they extract the same amount of resources: no matter how large their effects on aggregate output are, the difference in the reduction of aggregate output in these two cases is smaller than 1 percentage point.

This finding may have important implications for empirical studies on misallocation. It suggests that contrary to what is commonly believed in the literature, cross-country differences in firm size distributions or the correlation between distortions and firm size or productivity do not tell us much about the distortionary effects on aggregate productivity and output. In contrast, the amount of resources extracted seems to be more informative of the aggregate effects of distortions.

1.5.2 Restrictions on Establishment Creation

In this subsection, I use the calibrated model to study restrictions on establishment creation by multi-establishment firms. As mentioned above, there are policy distortions that restrict the establishment creation by multi-establishment firms, such as geographic restrictions on the US banking industry before the 1990s, which limited bank's ability to choose branch locations, and the restrictions on the entry of large international retail chains in India's retail industry. Since it explicitly models multi-establishment operations, my model is well suited to studying these restrictions.

In the benchmark economy, the cost of operating an additional establishment

	Benchmark						
Level (κ)	1	1	0.927	1	0.870	1	0.826
Size Dependency (ρ)	0	0.02	0	0.04	0	0.06	0
Statistics							
Aggregate Output	100	97.2	96.5	94.5	93.7	91.9	91.4
Wage	100	90.6	89.5	82.7	81.5	76.2	75.5
Mean Firm Size	100	73.4	99.9	56.2	99.6	43.7	99.7
Mean Establishment Size	100	76.4	99.8	60.6	99.4	48.9	99.5
Establishments Per Firm	100	96.0	100.1	92.5	100.2	89.6	100.2
Employment Share of Firms 500+	50.0%	42.0%	49.6%	35.3%	49.6%	29.1%	49.6%

† This table shows the effects on displayed variables associated to size-dependent and size-independent distortions at establishment level across steady states. Column 2 reports the values of displayed variables in the benchmark economy, most of which are normalized to 100. Column 3, 5 and 7 report the effects associated to distortions with different degree of size dependency, while Column 4, 6 and 8 report the effects associated to the corresponding size-independent distortions that extract the same amount of resources.

Table 1.7: Effects of Establishment-Level Distortions: Size-Dependent vs. Size-Independent

is τ_e unit of labor, which may capture the regulation of establishment entry, the communication cost between headquarters and affiliate establishments and so forth. Therefore, I model the restrictions on establishment creation as an increase in τ_e. I study the effects of restrictions that apply to a small sector, in which case wage w is determined exogenously and will not be affected by the restrictions, as well as restrictions that apply to the whole economy, in which case wage w is determined endogenously and will be affected by the restrictions.

Table (8) summarizes the cross-sectional and the aggregate effects of restrictions on establishment creation that raise τ_e by 100%. In both cases of fixed and endogenous wages, restrictions on establishment creation have large effects on the firm size distribution. The intuition is follows. As τ_e goes up, the cost of multi-establishment operations increases, and multi-establishment firms choose to operate a smaller number of establishments with larger employment sizes. Meanwhile, the increase in τ_e affects only multi-establishment firms, which are at the same time large firms, therefore both mean firm size and the share of large firms in the economy would decrease. In the case of fixed wage, a 100% increase in τ_e causes a 40% decrease in mean firm size and 17% decrease in the mean number of establishments per firm, and the employment share of firms with more than 500 employees drops from 50% to 28%. In the case of endogenous wage, a 100% increase in τ_e causes a 24% decrease in mean firm size and 15% decrease in the mean number of establishments per firm, and the employment share of firms with more than 500 employees drops from 50% to 33%.

In contrast, restrictions on establishment creation have very different effects on output in the two cases. In the case of fixed wage, total output in the sector where the restrictions apply to would drop by 33% after τ_e rises by 100%. However, in the case that restrictions apply to the whole economy and wage would affected by the restrictions, the effect of restrictions on aggregate output is much smaller: aggregate output only drops by 2% after τ_e rises by 100%. The reason is that, in the case of endogenous wage, equilibrium wage drops after the introduction of restrictions, which induces more firm entry. This in turn offsets the distortionary effects on aggregate output. This result confirms the important insights from Atkeson and Burstein (2010) in a different setting: firm entry plays an important role in offsetting the aggregate effects of incumbent firm's decisions.

	Benchmark	Wage is fixed τ_e doubles	Wage is endogenous τ_e doubles
Statistics			
Aggregate Output	100	62.7	97.7
Mean Firm Size	100	60.4	76.1
Mean Establishment Size	100	72.5	89.6
Total Number of Firms	100	100	131.4
Total Number of Establishments	100	83.3	111.6
Establishments Per Firm	100	83.2	84.7
Employment Share of Firms 500+	50.0%	28.0%	32.8%

† This table shows the effects on displayed variables associated to restrictions on establishment creation across steady states. Column 2 reports the values of displayed variables in the benchmark economy, most of which are normalized to 100. Column 3 reports the effects of restrictions on establishment creation when they apply to a small sector and double the cost of operating an additional establishment τ_e, in which case wage is fixed, and Column 4 reports the effects of restrictions when they apply to the whole economy and double τ_e, in which case wage changes endogenously.

Table 1.8: Effects of Restrictions on Establishment Creation

1.5.3 Discussion of Results and Related Literature

This paper contributes to a growing literature on misallocation and productivity, see Restuccia and Rogerson (2013) for a review of the recent literature. The most closely related paper is Hopenhayn (2014), which establishes theoretically the mapping between distortions and aggregate productivity, and constructs an employment-weighted measure of distortions. Contrary to what is commonly believed in the literature, he finds size distributions may not be informative of effects on aggregate productivity, and distortions that are correlated with firm size or productivity are not necessarily more damaging compared with other distortions.

My model extends the simple static framework in Hopenhayn (2014) along several important dimensions, but the results in my paper confirm the important insights in his paper. Moreover, the construction of the measure of distortions in Hopenhayn (2014) requires the information on firm size distribution in the undistorted and distorted economies, while the quantitative results in my paper suggest the amount of resources extracted via the distortions may summarize the aggregate effects of distortions to a first-order approximation.

Atkeson and Burstein (2010) build a general equilibrium model in which firms' decisions respond to a change in trade costs, and they find the change in trade cost can have large impact on firms' decisions, but the impact is not important for aggregate productivity. They emphasize the importance of firm entry (or product innovation), which largely offsets the effects of incumbent firm's decisions on aggregate productivity. This paper studies a different question, but the key results are similar in spirit: distortions may have vastly different effects on individual firms' decisions and firm size distribution, but once we control for total tax revenue, their effects on aggregate output are similar. And firm entry plays an important role in offsetting the aggregate effects of incumbent firm's decisions in this paper as well.

This paper is also related to studies on multi-national firms and foreign direct investment (FDI) in the trade literature. Markusen (1984) builds an equilibrium model of multi-national firms based on knowledge capital, i.e., intangible assets that have a joint-input feature, which give rise to the economies of multi-plant operations; see Markusen (1995) for a review of subsequent contributions to the literature. McGrattan and Prescott (2009) incorporate technology capital into neoclassic growth model and use it to quantify the gains from opening to FDI. Ramondo (2014) builds a multi-national firm model that combines Lucas (1978) with Eaton and Kortum (2002), and use it to quantify the gains from opening to FDI.

Quantitative studies in the literature find large gains from opening to FDI, while in a domestic setting, my paper finds relatively small losses to aggregate productivity from restrictions on establishment creation. One possible explanation for the relatively small losses is, labor is allowed to move freely in my model, while there are severe restrictions on labor mobility across national boundaries in the trade literature. In this sense, the losses from restrictions on establishment creation in this model can be regarded as a lower bound, and it would be interesting to see how restrictions on labor mobility would change the results.

1.6 Conclusions

I build a multi-establishment firm model and use it to study distortions that cause misallocation among establishments and firms. I find that size-dependent distortions to establishments and to firms have very different effects on firm size distribution, but have similar effects on aggregate output. I also use the model to study restrictions on establishment creation, and find they have large effects on output when applying to a small sector, and small effects on aggregate output when applying to the whole economy.

This paper focuses on the steady-state effects of misallocation. However, the analytical framework of this paper can be extended to study important questions regarding productivity growth. Unlike Europe, the US experienced a surge in productivity growth in the 1990s. Meanwhile, important sectors such as the banking and retail trade underwent a massive reallocation of resources from single-establishment to multi-establishment firms. The model in this paper can be modified to study how this massive reallocation and its interaction with information technology affect productivity growth. This should contribute to our understanding of the difference in productivity growth between the US and Europe.

It would also be interesting to extend the model to multi-country settings, and use it to understand the role of multi-national firms in global production and assess the gains from openness to trade and foreign direct investment.

Chapter 2

THE SIZE DISTRIBUTION OF FIRMS AND INDUSTRIAL POLLUTION

2.1 Introduction

Rapid economic growth in China has successfully pulled hundreds of millions of people out of poverty in the last few decades. However, as both the pace of industrialization and urbanization accelerate, the public has been increasingly concerned with the environmental consequences of economic growth. [1] In recent years, major metropolitan cities across the country suffered from atmospheric haze pollution. A large number of people have been affected by frequent incidents of emergent and cumulative water contamination. [2] As stated in the *Report on the Work of the Government 2014*, the Chinese government has vowed to undertake a campaign to fight against environmental pollution. For this purpose, the State Council has allocated almost \$600 billion of special funds for controlling air and water pollution. [3] The primary target of the campaign is to reduce industrial pollution. In order to provide effective policy prescriptions, the key question is then what is the driving force behind the heavy industrial pollution by Chinese manufacturing firms?

We show that firm size is an important factor in explaining the high industrial pollution emission problem in China. There are two observations that motivate our

[1] For general surveys of the current situation of China's environmental pollution see Vennemo *et al.* (2009), Zheng and Kahn (2013) and the references therein. For media press coverage, see for example the Symposium "Choking on growth — Examining the Impact of China's Epic Pollution Crisis" in *The New York Times* in late 2007.

[2] A wealth of literature has since investigated the causal relationship between pollution and various aspects of human's well-being. For health consequences of water pollution, see for example Economy (2004), Ebenstein (2012), Zhang (2012), and Yang and Zhuang (2014). Graff Zivin and Neidell (2013) provides an excellent survey of the related literature.

[3] The funds are CNY 1.6 trillion (\$260 billion) under *Air Pollution Prevention and Control Plan* and CNY 2 trillion (\$320 billion) under *Water Pollution Prevention and Control Plan.*

inquiry into firm size:

(i) Small firms have much higher pollution intensity (pollutants discharged per unit of production) than large firms, indicating that firm size differences can potentially have a large effect on measured aggregate industrial pollution. Using firm-level emission data on water pollution from *the First National General Survey of Pollution Sources*, we find 7- to 32-fold differences in pollution intensity between firms with total output in the top and bottom quartiles for the top 5 polluting industries in China. [4] Furthermore, we find that large firms pollute less because they use production technologies that are more environmentally friendly, and pollutant treatment technologies that are more advanced. Put differently, large firms not only generate less pollutants, but also remove a larger proportion of them from the discharged wastewater.

(ii) Small firms account for a larger fraction of production in manufacturing sectors in China than in the U.S. Using data from *the First China National Economic Census* and *the Statistics of U.S Businesses*, we document that for the top-5 polluting industries in China, firms with more than 400 employees account for 40% of the total employment, while for their American counterparts the number is close to 70%. [5]

We investigate the role of product market frictions in shaping the firm size distribution in China and subsequently assess quantitatively their impact on aggregate industrial pollution and other macro aggregates, as well as their interaction with en-

[4]OLS regression indicates that as the total sales increase by 1%, the total emission increases by 0.62%. Studies using either data from other countries or from a selective group of Chinese firms also point to the negative correlation between firm size and pollution intensity. See, for instance Shapiro and Walker (2015) for the U.S; Dasgupta *et al.* (1998) for Brazil and Mexico; and Bloom *et al.* (2010) for energy use in the UK.

[5]See Axtell (2001), Luttmer (2007) and Rossi-Hansberg and Wright (2007) for theories and evidence regarding the heavy right tail of U.S. firm distribution.

vironmental policies. We use wedges of average factor product to measure implicitly the level of product market frictions [Hsieh and Klenow (2009) and Restuccia and Rogerson (2013)]. Examples of such product market frictions include local protectionism and trade barriers which impede the inter-regional flow of goods, as well as various administrative costs, etc. We find that product market frictions affect disproportionately large productive firms, which limits the expansion of these productive firms. Therefore, unproductive firms are allowed to survive, which results in output loss. Since the adoption of advanced pollution treatment technology requires fixed installation costs, product market frictions affect aggregate pollution via two channels. First, less firms are willing to install clean technologies because firms earn less profits in a market with frictions. Second, firms using clean technologies account for a smaller market share. As a result, output is lower *and* pollution level is higher.

To guide our analysis, we use a neoclassical growth model with heterogeneous production units [the Lucas (1978) span-of-control model] featuring product market frictions, imperfect environmental monitoring and endogenous pollution treatment technology choice (clean and dirty). In our model, there is a stand-in representative household with a continuum of members that are endowed with different managerial talents. Household members make occupational choice decisions based on their talents before entering the economy. A new feature of our model is that upon entering the market, the entrepreneurs have to make decisions on which treatment technology to install. In reality, there are two stages that firms can take actions to cut their emission level. Firms can reduce the total quantity of pollutants generated during the production stage by using environmentally friendly production technologies, or reduce the end-of-pipe emission by adopting more advanced treatment equipments for a given amount of pollutants generated. In this paper, we focus mainly on firm's choice of treatment technology. We capture the decrease of pollution intensity during the production stage in a reduced-form way, which is calibrated to data. The installation of clean technology requires some fixed costs, however, firms with clean technology will not be punished by the environmental agencies. The fixed costs associated with clean technology lead to increasing returns to scale, which implies that in our model, only large firms find it profitable to install clean technology. The optimal scale of operation of a firm is determined by the managerial talent of the entrepreneur and the frictions that the firm faces. Therefore, for a given distribution of managerial talents and market frictions, the model generates endogenously the distributions of firm sizes and employment, and that of the clean technology adoption.

To discipline our quantitative analysis, we require that our benchmark model with product market frictions matches the firm size and employment distributions observed in China. We then conduct two policy experiments: in the first one we eliminate the product market frictions completely, while in the second we increase the regulation such that the fraction of firms adopting clean treatment technology is the same as in the first experiment. We subsequently calculate and compare measures of output, consumption, productivity and aggregate pollution under the two policies.

Our quantitative results show that elimination of product market frictions increases output by 30%, increases the fraction of firms using clean technology by 27% and decreases pollution by 20%. The improvement in output is expected, since the elimination of frictions allows the production resources to be allocated more efficiently in the model. The drop in pollution comes from both the reduction in pollutants generated during the production stage, and the increase in adoption rate of clean treatment technologies at the treatment stage. Each stage contributes to about 50% of the total reduction. The expansion of productive firms is key to both channels. On the other hand, environmental policy which increases the expected punishment reduces pollution by only about 10% and has very little effect on output. Moreover, we find that the environmental policy improves resource allocation on the *extensive* margin by driving small unproductive firms out of the economy. However, the allocation worsens at the *intensive* margin in the sense that among the remaining active firms, the production of medium sized firms expands more at the expense of large firms.

To assess quantitatively the importance of the size-dependency of these frictions, we solve a version of the model where all firms in the economy face the same level of frictions. In our model, the size-dependency of the frictions does not imply large output loss. However, the size-dependency of the frictions assumes a central role in determining the pollution level. This finding is consistent with Hopenhayn (2014). The author shows that it is the total amount of resources that are affected that determines the effects of the frictions as opposed to who is affected. Our results complement his findings by demonstrating that while the size-dependency of the frictions does not affect aggregate output by too much, its effect on the pollution level is much larger.

Our findings call for a change in the policies that address the pollution problem pairing urbanization and industrialization. The GDP-oriented promotion scheme in China, under which whether a local government official is promoted depends on the growth rate his/her governing region, has been identified as the source of industrial pollution problem [Jia (2014)]. In this paper we emphasize on the roles of product market frictions and firm size distribution. From a policy perspective, our results suggest that a double-dividend where increase in the output and reduction in pollution are achieved simultaneously [Goulder (1994)] can be attained with policies that target at reducing the economic frictions which prevent talented entrepreneurs from operating their business at the scale that is necessary for the adoption of clean technologies. In fact, one implication of our results is that the GDP-oriented promotion scheme does not necessarily yield increased environmental pollution. In this way, we view the insight provided in our paper as a complement to those in the political economy literature.

Related Literature.—Our paper contributes to four strands of literature. First, we contribute to a broad literature analyzing the environmental consequences of economic activities. Modern discussions in this area are usually traced back to the seminal work by Grossman and Krueger (1993, 1995) where the authors documented an inverse U-shaped relationship between various metrics of pollution levels and output per capita. This relationship, due to its resemblance of the famous Kuznets curve [Kuznets (1955)], is thus referred to as the *environmental Kuznets Curve* (EKC) in subsequent literature. The international trade community has devoted considerable efforts to studying the underlying economic mechanisms of the EKC. Copeland and Taylor (2004) provide a thorough and complete survey of the early contributions. Most of those studies focused on decomposing the pollution to the scale, technology and industry composition effects using reduced-form methods. There is however, a very recent growing literature on CO_2 emissions in the trade community where heterogeneous firms models are involved theoretically [Barrows and Ollivier (2014) and Shapiro and Walker (2015)]. We view our approach as complementary to that work. In this paper, we build a macroeconomic model with rich quantitative implications which facilitates the investigation of policy related questions. Our focus on the firm size distribution and treatment technology adoption also distinguishes our paper from that literature, which studies the roles of product mix, consumer preference, etc. In connection with literature focusing on the cross-country comparison of firm size distributions [for example Poschke (2015)], our paper is also a candidate of structural interpretations of the EKC.

Second, there is also a growing literature studying the causes of China's industrial pollution. In accord with the views widely covered by the media, most these studies have identified political factors like the GDP-oriented promotion schemes [Wang *et al.* (2008), Jia (2014)] and differential policy treatment to firms' with different ownership rights structures [Jiang *et al.* (2014)] as important factors that cause massive industrial pollutant discharge. In this paper, we argue that at the micro level, the effects of economic frictions also play a major role. Political factors may act as amplification of the effects of economic frictions, and vice versa. Policy prescriptions aiming at reducing such economic frictions could potentially overcome the problems of poor policy implementations and quick rebound that are constantly disturbing policy makers. Another difference between our paper and these studies is that we have access to a universal coverage database which contains information on *directly observed* firm-level discharge of various pollutant.

Third, our paper is closely related to an important and growing literature on the misallocation of resources across heterogeneous production units and its implications on macroeconomic aggregates. The early contributions are seminal work from Hopenhayn (1992) and Hopenhayn and Rogerson (1993). [6] Our paper relates particularly to Guner *et al.* (2008) and Adamopoulos and Restuccia (2014) where the role of size-dependent policies is examined. We contribute to this literature in two ways. First, we provide empirical evidence on the potential role of product market frictions in generating differences in size distributions between China and the U.S, using the *indirect* approach by Hsieh and Klenow (2009, 2014). Second, we extend the discussion of implications of size distribution on aggregate output and TFP to aggregate pollution.

Lastly, our paper also connects to the literature on technology adoption in macroe-

[6] See Restuccia and Rogerson (2008), Guner *et al.* (2008), Hsieh and Klenow (2009) and Adamopoulos and Restuccia (2014) for recent discussions amongst others.

conomics. The seminal work by Parente and Prescott (1994) introduces frictions in technology adoption as a candidate for generating the cross-country productivity difference. A number of papers were dedicated to understanding technology diffusion since then. This paper investigates in particular the role of product market frictions and size distribution in impeding the adoption of clean production technology. On this aspect, our paper inherits the intuitions from two early studies in economic history about tractors in the U.S—Clarke (1991) with the role of market frictions and Olmstead and Rhode (2001) with the role of size distribution. We also view our study as complementary to the paper by Acemoglu *et al.* (2012). There the authors analyze the optimal policies to promote the advancing of clean production technology, while our paper answers the related question of under what circumstances will these newly invented technologies eventually be installed by firms.

The rest of the paper proceeds as follows. The next section documents facts pertaining to pollution intensity differences across firms and the comparison of firm-size distributions between China and the U.S. We describe the model in Section 2 and calibrate its benchmark version in Section 3. In Section 4 we perform several policy experiments to study the interaction between product market frictions and environmental policies. We conclude in Section 5.

2.2 Empirical Evidence

In this section, we document the key empirical findings regarding the size-intensity relationship and the comparison of size distributions between China and the U.S that motivate our study. We start with a brief introduction of the data that we use. We then move on to explain the empirical findings. Using an accounting exercise, in the last section, we show that size distribution has a sizable effect on aggregate pollution.

2.2.1 Data Sources

There are three major data sources that we draw upon in this paper: (i) the First National General Survey of Pollution Sources, (ii) the First China National Economic Census and (iii) the Statistics of U.S. Businesses. These three data sources are used to calculate the pollution intensity of Chinese manufacturing firms and the size and employment distributions of manufacturing firms in China and in the U.S. They are referred to in the remainder of this paper respectively by their acronyms NGSPS, CNEC and SUSB. [7]

National General Survey of Pollution Sources.—The NGSPS is a joint effort of multiple national ministries in China. The survey records data for year 2007. It is designed to cover all entities and self-employed households which emit pollutants in China. The complete survey consists of four components: industrial pollution sources, agricultural pollution sources, domestic pollution sources and facilities for centralized treatment of pollution. For the purpose of this paper, we use only information from the industrial pollution sources. The variables we use are: the quantity of major pollutant generated and discharged, the total value of production, the book value and the annual operating costs of pollutant treatment equipments, the type of treatment equipments, the firm's industry (four-digit GB/T4574-2002), the ownership classification, and the province. Specifically, the NGSPS contains information on discharges of air and water pollutant and solid waste. Here we focus on water pollution because the data are more accurately measured. However, we expect that the main results of this paper could be applied to other pollution sources as well. The raw data contain 921,004 firms.

[7] In the interest of space, we leave more detailed description of these data, variables and sample selections criteria to the online Appendix. The online appendix is currently under completion, results are available upon request. Please address all correspondence to zjutangxin@gmail.com.

China National Economic Census.—The CNEC is conducted by the National Bureau of Statistics (NBS, henceforth) in year 2004. It is designed to cover all legal entities, industrial entities and privately-owned businesses which undertake economic activities in secondary and tertiary industries in China. For our purpose we use observations which belong to the manufacturing sector. The variables we use are: the total value of production, the labor compensation, the book value of capital stock, the number of employees, the firm's industry (four-digit GB/T4574-2002), the ownership classification, and the province. [8] The number of firms covered in NGSPS and CNEC, 921,004 thousand and 1,375,148 million are broadly consistent given that NGSPS further requires that a production entity to have pollution sources in order to be included. [9]

Statistics of U.S. Businesses.—The SUSB is conducted by the U.S Census Bureau and is an annual series that provides national and subnational data on the distribution of economic data by enterprise size and industry. It contains the number of firms and total employment by sector (up to six-digit 2002 NAICS) and enterprise size groups which we use.

2.2.2 Firm-level Pollution Intensity

Sample Selection.—It is well established in environmental science that industrial waste is typically concentrated in a handful of sectors. Even within narrowly defined

[8]We emphasize here that it is important that we use the CNEC rather than the *Annual Surveys of Industrial Production* for which data of year 2007 is available (the same year that the NGSPS covers). The reason is that CNEC surveys firms of all sizes as opposed to only firms with a revenue of more than CNY 5 million in the case of the annual surveys. In 2004, the number of firms and employees covered by the annual survey are respectively 276,410 and 66,725,059 while those covered in the census are 1,375,148 and 93,541,923. Therefore we would be missing 28.6% employment and 79.9% firms had we used only the annual survey. However, the basic features like variable definitions are essentially the same in these two datasets. Therefore we would like to refer interested readers to Brandt *et al.* (2012) which contains a detailed description of the annual surveys.

[9]More detailed comparisons on the statistical features of NGSPS and CNEC are available upon requests.

manufacturing sectors, pollutant emissions are usually concentrated among firms that engage in some particular manufacturing processes. To address this issue, the NGSPS divides the complete sample into two large groups—*key sources* and *regular sources*—where firms identified as "key sources" are those that are most polluting. [10] We focus on the key firms and use *Chemical Oxygen Demand* (COD, henceforth) as an example in the main text. COD is the amount of oxygen consumed when a chemical oxidant is added to a sample of water. It is an indirect measure indicating the overall quantity of contaminants that will eventually cause oxygen loss and thus death of living creatures. Table 2.1 lists the percentage of key and regular firms that have positive emissions of different pollutants. We focus on key firms because the quality of the data of these firms are higher, and most regular firms emit very little pollutants. We choose COD because it allows us to keep most observations from the data. Other pollutants are discharged by significantly less number of firms which raises sample selection concerns. Moreover, COD emission is highly correlated with the emission of other pollutants. [11] Finally, we focus on the *measured* end-of-pipe discharges and present results for the top-5 polluting industries. Altogether, this leaves us with 29,019 firms.

Table 2.2 contains basic statistics about these industries. We see from it that the key firms in the top-5 polluting industries are fairly representative of China's industrial pollution situation: in particular, these industries combined contribute to 77% of the total industrial COD emission (Figure 2.1 provides a graphical illustration); the key firms are responsible on average for more than 90% of the within sector emission and for more than 80% of the within sector output.

[10] The Online Appendix contains a detailed description of the definition of the key sources.

[11] Take the Paper and Paper Product industry for example, the correlation between the emission of COD and that of NH_4^+, $\mathrm{corr}(\mathrm{COD}, \mathrm{NH}_4^+) = 0.82$, and that between COD and BOD, $\mathrm{corr}(\mathrm{COD}, \mathrm{BOD}) = 0.94$.

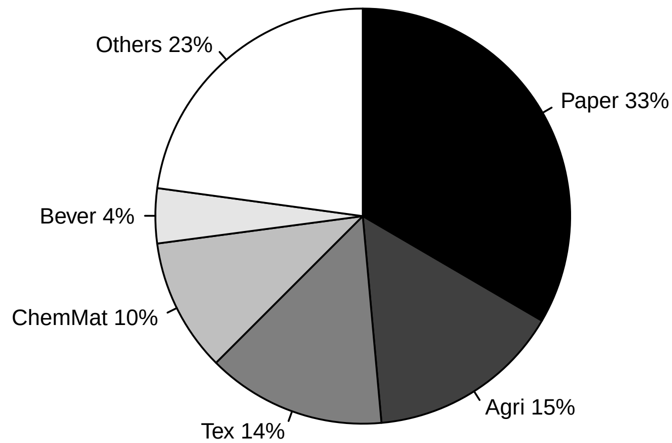

Figure 2.1: COD Emission by Sectors

	Waste	COD	Petro	NH_4^+	BOD	CN	Cr^{6+}	Phenol	As	Cr	Total
Key	76.2	73.2	31.4	25.2	17.5	4.90	4.86	2.42	2.27	2.01	106,067
Reg	35.2	28.3	7.91	6.49	2.56	0.13	N/A	0.04	0.07	N/A	814,937

† Data Source: National General Survey of Pollution Sources. The acronyms are respectively referring to: Wastewater, Chemical Oxygen Demand, Petrochemicals, Ammonian, Biochemical Oxygen Demand, Cyanidium, Hexavalent Chromium, Volatile Phenols, Arsenium and Chromium.

Table 2.1: Percentage of Firms with Positive Emission by Pollutants

Pollution Intensity and Firm Size.—We define pollution intensity per unit value of production as

$$\text{Intensity} = \frac{\text{Total COD Emission}}{\text{Total Value of Production}}.$$

We group the firms into quartiles based on their total value of output. For each industry, we calculate the output-weighted average of pollution intensity of the firms in each quartile. Table 2.3 reports the results. For the Paper and Paper Product industry, the pollutants emitted per unit of production by the firms in the bottom quartile is 6.7 times of that by the firms in the top quartile. The difference can be as

	Paper	Agri	Tex	Chem	Bever	Med	Fer	Petro	Food	Fib
Frac[a]	33.4	15.2	14.0	10.4	4.27	2.98	2.49	2.32	2.30	2.15
% Emi[b]	99.6	91.8	91.1	99.7	65.1	92.9	99.9	99.9	96.4	97.8
% Pro[c]	87.2	69.3	48.3	98.6	88.1	95.7	99.3	99.7	98.5	91.9

† Data Source: National General Survey of Pollution Sources. The acronyms are respectively referring to (with two-digit GB/T4547-2002 classification code in the parentheses): Paper and Paper Products (C22); Processing of Food from Agricultural Products (C13); Textile (C17); Raw Chemical Materials and Chemical Products (C26); Beverages (C15); Medicines (C27); Mining, Smelting and Pressing of Ferrous Metals (C32); Processing of Petroleum, Coking, Processing of Nuclear Fuel (C25); Foods (C14); Chemical Fibers (C28).

[a] Relative contribution to total COD emissions by sectors.

[b] Percentage of total COD emissions accounted for by key firms.

[c] Percentage of total production accounted for by key firms.

Table 2.2: Statistics of Top-10 Polluting Industries by COD

large as 31.4 times, as is the case of the Beverage Manufacturing industry. Moreover, the pollution intensity decreases continuously as the size of the firm becomes larger. This can also be seen from the scatter plot of the logarithm of intensity against that of the total value of production. We plot the Paper and Paper Product industry in Figure 2.2 for an example, scatter plots for the other four industries are left in the Online Appendix. A significant negative correlation between log-intensity and log-production in the data can be seen in Figure 2.2.

To further examine the statistical property of the relationship between intensity and production level, we regress the log-emission on the log-sales, including a complete set of dummies for two-digit industry ($\mathbf{X_s}$), province ($\mathbf{X_p}$), and ownership rights ($\mathbf{X_o}$):

$$\log(\text{COD}_i) = \underset{(0.37)}{-3.36} + \underset{(0.01)}{0.62} \times \log(\text{Sales}_i) + \mathbf{X_s}\gamma_1 + \mathbf{X_p}\gamma_2 + \mathbf{X_o}\gamma_3 + \varepsilon_i. \tag{2.1}$$

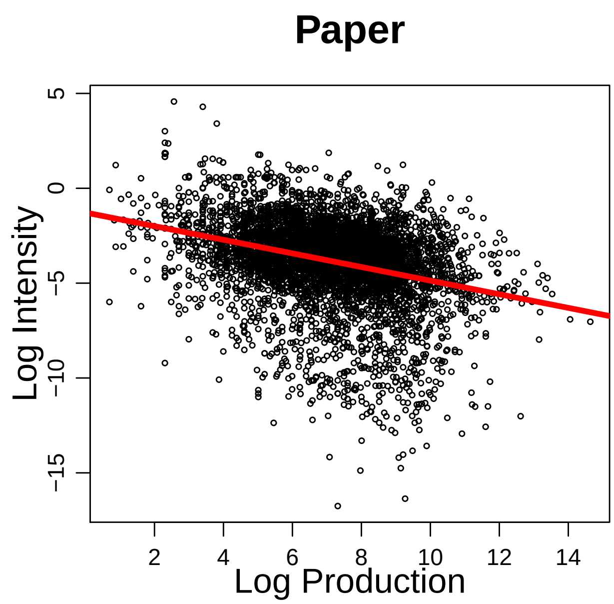

Figure 2.2: Pollution Intensity against Production

Source: National General Survey of Pollution Sources. Line: Least square fit.

The estimates are all statistical significant at 0.1% level with the standard errors reported in the parentheses below the estimates. [12] The estimate implies that as the total sales increases by 1%, the total emission increases by less than 1%, in particular by 0.62% here. This suggests that the emission intensity is decreasing as the sales of the firm increases. More specifically, by subtracting $\log(\text{Sales}_i)$ on both sides of Equation (2.1), the elasticity between pollution intensity and total sales is -0.38, which means that other things equal, a 1% increase in the total sales is associated with a 0.37% decrease in pollution intensity. The estimation has a R^2 of 0.55, which suggest that a fair amount of variation can be explained by variations in the total

[12] We have also estimated the relationship using other econometric specifications. For instance, we estimated versions of Equation (2.1) for each industry, and with robust standard errors clustered on different groups. All the regressions suggest the same negative relationship between intensity and size qualitatively. The estimation results of the other specifications, as well as interpretations of the coefficients before the dummies are included in the online Appendix.

	Quartile of Firm Sales			
Industry	QU_1	QU_2	QU_3	QU_4
Paper	6.7	3.2	2.0	1.0
Food Processing	20.8	7.6	3.6	1.0
Textile	8.3	3.6	2.4	1.0
Chemical Materials	6.7	3.8	2.7	1.0
Beverage	31.4	18.7	4.7	1.0

† Data Source: National General Survey of Pollution Sources. QU_1 to QU_4 represent respectively the bottom to the top quartile. The pollution intensity of the top quartile of each industry is normalized to one.

Table 2.3: Pollution Intensity and Production Level

sales and in the three sets of dummies.

2.2.3 Firm Size and Treatment Technologies

The negative size-intensity relationship we document in Section II.B does not explain why larger firms pollute with less intensity. To answer this question, we exploit the detailed information in the NGSPS on the *end-of-pipe* wastewater treatment equipment that firms use. The NGSPS groups wastewater treatment technologies in five categories: *physical, chemical, physio-chemical, biological* and *combined* technologies. In the subsequent analysis, we drop physio-chemical technologies because less than 0.5% firms adopt this type of equipments. The combined technologies are different combinations of biological technology with other technologies. They demonstrate very similar features as biological technologies. We therefore group them with biological technologies. Several examples of the actual technologies attributed to the three base categories (physical, chemical and biological) are provided below:

(i) Physical: Filtering, Centrifuging, Precipitation Separation, etc.

Technology	25%	Median	75%	Mean	Adoption Rates
Physical	39.54%	77.81%	87.83%	63.37%	25.79%
Chemical	74.96%	81.29%	86.78%	69.77%	34.50%
Biological	78.87%	86.77%	91.27%	80.90%	39.71%

† Note: The numbers reported are for the Paper and Paper Product (C22) industry. Treatment Efficiencies is defined as $1 -$ COD Emitted/COD Generated.

Table 2.4: Treatment Efficiencies and Adoption Rates

(ii) Chemical: Oxidation-reduction, neutralization, etc.

(iii) Biological: Aerobic Biological Treatment, Activated Sludge Process, etc.

We are interested in the following features of these technologies: processing efficiency, designed processing capacity and installation costs. [13] To control for potential heterogeneities in production processes across different industries, we use the Paper and Paper Product industry (C22) as an example. [14] We proxy the processing efficiency using one minus the ratio of emitted COD to generated COD. Table 2.4 shows the quartiles of processing efficiencies and percentage of firms adopting each type of technology. The designed processing capacity (in tons) and actual installation costs (in 2007 CNY) that are needed for the equipment to function properly can be retrieved directly from the data. We calculate the unit capacity cost (or average cost of capacity) by dividing the installation cost by capacity. In Figure 2.3, in clockwise order we plot the density functions of processing capacity, installation

[13] We do not include the annual operating costs in our analysis because on average, the ratio of operating costs of the treatment equipments on the annual value of production is about 1.5%. Furthermore, the median of this ratio is less than 0.5%, suggesting that operating costs are almost negligible for more than 50% of the firms. Therefore, the operating costs alone is unlikely to affect firm's treatment technology adoption decision. Adding the operating costs to the installation costs will not change the results.

[14] Pooling all polluting industries together yields very similar results and are hence left in the Online Appendix.

costs, total value of industrial output and unit capacity cost by technology type. For all panels, log-scale is used in the horizontal axes. Broadly speaking, biologi-

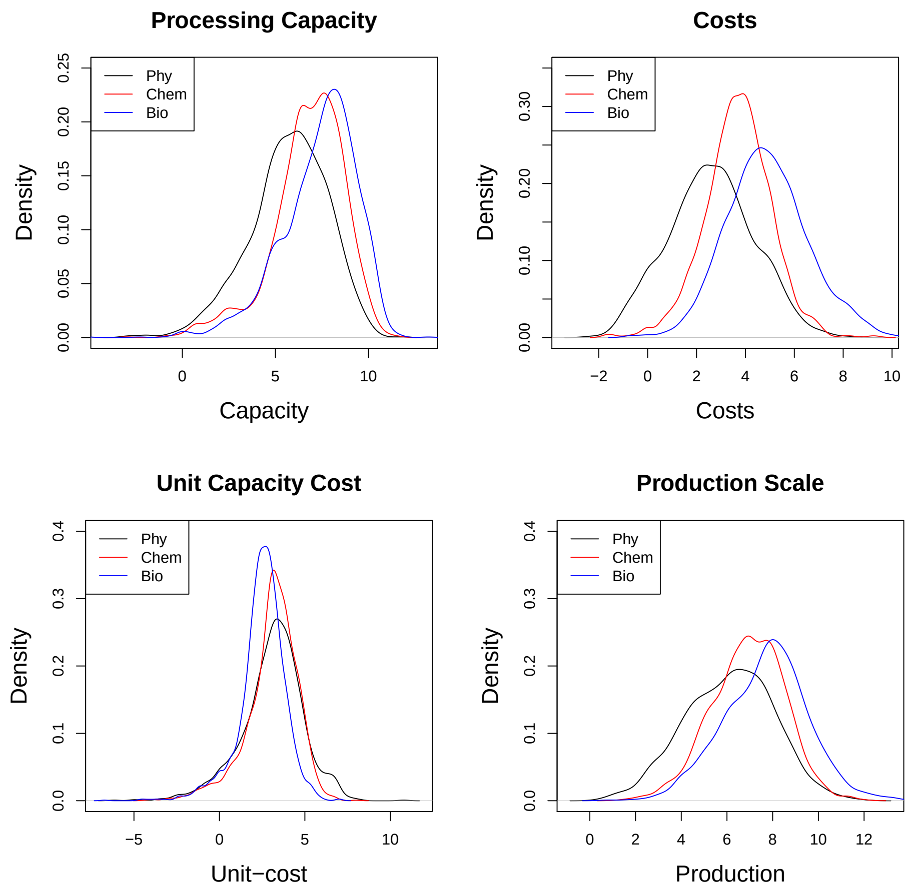

Figure 2.3: Technical Features of Different Treatment Technologies

Source: National General Survey of Pollution Sources. In all panels, the horizontal axes are in log-scale.

cal technologies have the best processing efficiency, the largest processing capacity, the highest installation costs but the lowest unit capacity cost. More specifically, the mean (median) processing efficiency of biological technology is 17% (10%) higher than the physical technology. The evidence points to a fixed costs type of mechanism behind the less pollution intensity by large firms. In particular, although biological technologies are more advanced in terms of processing capacity and efficiency, they are also more costly. Combined with the lower unit capacity cost that is displayed by the biological technology, the evidence can be rationalized by the existence of a fixed cost, which works essentially in the same way as the decreasing average cost when a fixed cost is involved. Such a modeling strategy will imply that small firms lack the profit margins that are needed to take advantage of the returns to scale exhibited by biological technologies, while at the same time, large firms are more likely to adopt these more advanced technologies.

Notice that the above results are all about the *end-of-pipe* treatment technologies, and we have made no statement about factors that could lead to less COD *generated*. In fact, in the data the COD generated per unit of production is also decreasing in total value of output. It is possible that more productive technologies are also more energy-efficient and environmentally friendly. An example from the *Handbook of Emission Coefficients* by the *Chinese Academy of Sciences* is as follows. Two technologies in paper pulp manufacturing use different inputs: bagasse and wood. While bagasse is used mostly by firms with annual production of less than 100 k-tons and with COD generation of 140-180 kg per ton, wood is used mostly by firms with annual production more than 100 k-tons and with COD generation of 30-55 kg per ton. Another example is from Bloom *et al.* (2010). The paper uses data of more than 300 manufacturing firms in UK and finds that better management practices are associated with both improved productivity and lower greenhouse gas emissions. In this paper, we focus on firms' decisions on treatment equipment adoption, and modeled the intensity reduction during the production stage exogenously.

2.2.4 Firm-Size Distributions

The negative correlation between pollution intensity and production scale implies that, ceteris paribus the relative contribution to total output by large and small firms could significantly affect the industrial pollution at the aggregate level. Therefore, for our purpose, it is pivotal to understand the industrial structure—the number, size and employment of firms—in China. To gauge our comparison, we look at size and employment data in the U.S as a yardstick. We choose the U.S to be a benchmark of the comparison for two reasons. First, it is generally agreed among macroeconomists that among all the economies in the world, the U.S economy is perhaps the closest counterpart to an undistorted market economy. Firms in many European countries are subject to various labor market restrictions and hence their distributional properties are less likely to be representative of a frictions-free benchmark. Second, China and U.S are both large economies with complete sets of industrial sectors. For disaggregated studies like ours, it is important that we find comparable counterparts in the benchmark country. Contrasting the industries in China with those in European advanced economies, it would be problematic to find comparable counterparts, or the corresponding industries are of significantly smaller production scale on aggregate.

The distribution of employment by firms of different sizes is the closest related concept to our analysis. Ultimately, what is crucial to the quantity of pollutants discharged is how much production is produced by small and large firms respectively, not how many of them there are in the economy. Employment is a good proxy for production share because it has been firmly established that labor compensation is strongly correlated with total production. Nevertheless, we provide evidence on both the size (the size distribution according to the number of firms) and employment (the size distribution according to employment shares) distributions in this section, in order to allow for comparisons with existing studies. We use the *International Standard Industrial Classification of All Economic Activities, Rev.3.1* (ISIC Rev 3.1) published by the United Nations to bridge different industrial classification systems adopted by China (GB/T4574-2002) and the U.S (NAICS 2002). More specifically, crosswalks of GB/2002 at four-digit level and those of NAICS/2002 at six-digit level are issued by China's NBS and the U.S Census Bureau. The results presented in this section are from matching at the disaggregated level (four-digit GB with six-digit NAICS). [15]

The SUSB organizes data according to enterprise size group rather than firm or establishment size which relates closer to what we want. In particular, for each size bin, the SUSB reports the total number of firms, establishments and employees summed up across all enterprises that fall in that size bin. We cannot use the size and share distribution of the enterprises because according to the definition in SUSB, a large enterprise could consist of firms and establishments in different locations, of different sizes and even in different sectors. However, the number of firms includes only those firms that are categorized as belonging to one particular industry. Therefore, we approximate the firm size distribution using the average firm size of a particular size group, which is calculated by dividing the total employment by the number of firms. We then assign groups of firms to different size bins according to their average size. Such imputation introduces approximation errors in a complex way. [16] However, we argue that most approximation errors lie in the upper tail of the distribution since it is less common for enterprises with less than 200 employees to have multiple firms

[15] Those from matching at a more aggregated level (two-digit GB with three-digit NAICS) yields very similar results

[16] Nevertheless, it is reassuring that the size and employment distribution calculated from the year 2000 data used in Rossi-Hansberg and Wright (2007), which is constructed from Census' micro-data (as opposed to imputed from the aggregated data here), give very similar results. These results can be found in the Online Appendix.

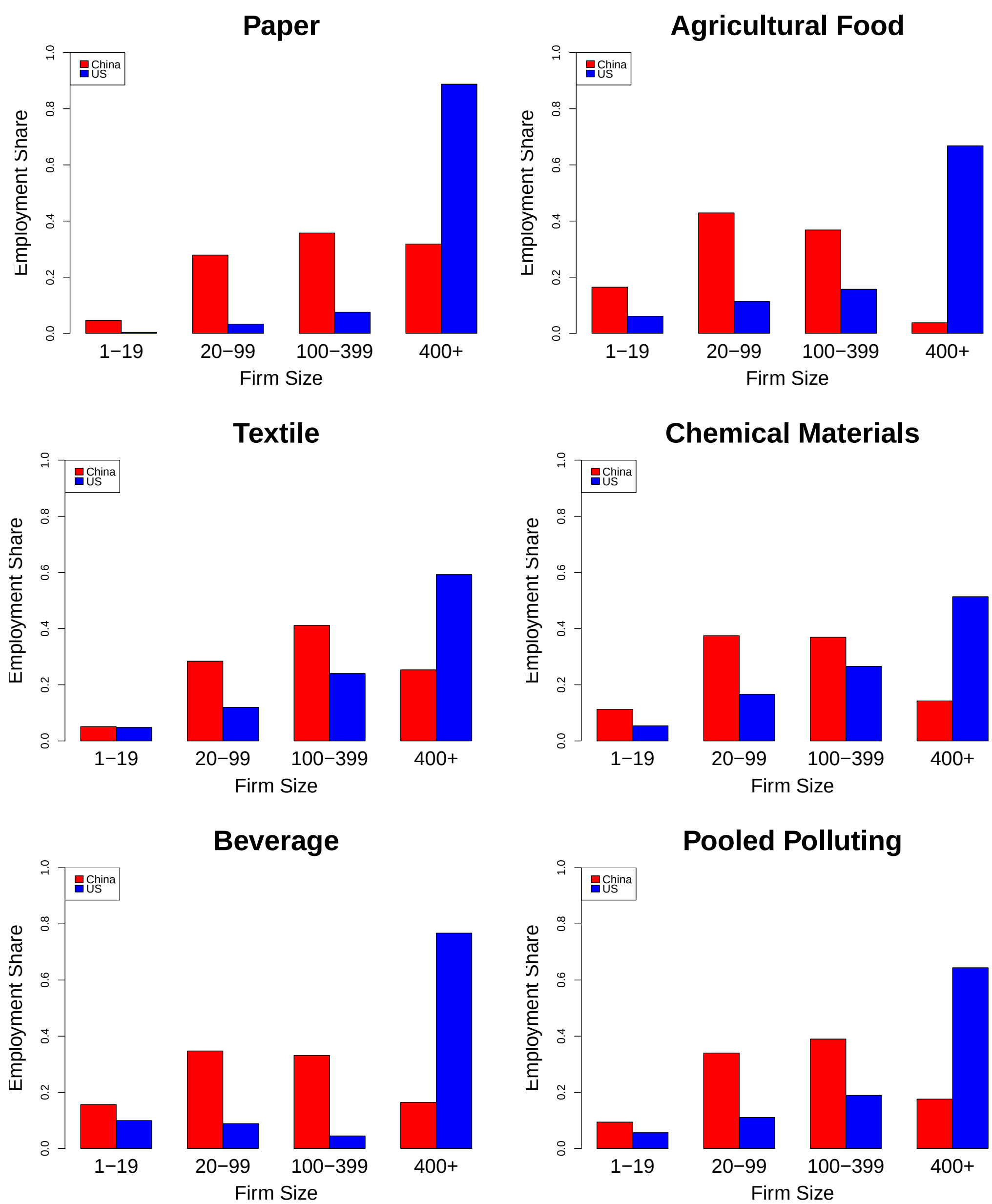

Figure 2.4: Employment Distribution

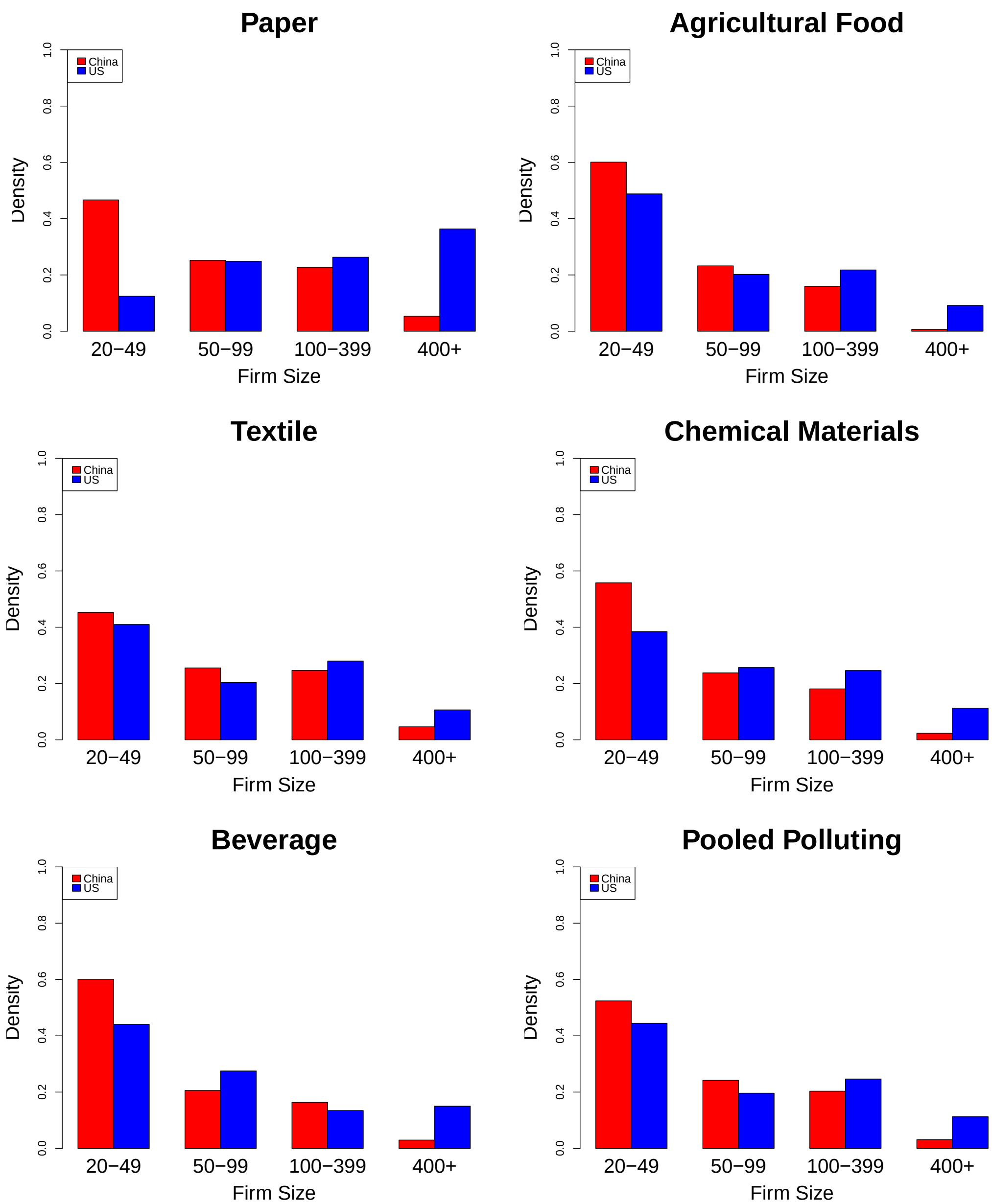

Figure 2.5: Firm Size Distribution

or establishments. To further reduce the approximation noise, we group the size bins into four main groups: 1–19, 20–99, 100–399 and 400+. We drop firms with less than 19 employees when plotting the size distribution because we think many of them may not be engaged in the actual production process but will significantly change the shape of the *size distribution*. Including or dropping these firms does *not* change the employment distribution because the calculation is essentially a weighted average with the number of employees as the weight. These small firms are thus weighted much less than the large ones. However, the firm size distributions are affected because of the sheer amount of these firms. [17]

The employment distributions for each of the top polluting industries and all industries pooled together are shown in Figure 2.4. Similarly, those of the firm size distributions are contained in Figure 2.5.

For all panels in Figure 2.4, we see that the share of employment of firms with more than 400 employees in the U.S is significantly higher than the one in China. More specifically, for the paper manufacturing industry, more than 90% of the workers in the U.S are hired by firms with more than 400 employees while in China, the number is less than 40%. Overall, looking at these industries together, approximately 70% employment is in the large firms in the U.S while in China the number is only 20%. [18] These features of the data are consistent with Wang and Whalley (2014) where the authors compare the manufacturing concentration ratio (the share of market occupied by the largest firms) between China and the U.S. According to Table 1 in their paper,

[17] We choose not to use establishment as the unit of our analysis because more noise is likely to be introduced by the approximation procedure that we adopt. However, results regarding the employment distributions, which are our ultimate interest here, is quite robust across variations. What is less robust is the firm size distribution, which is less relevant to our conclusions. All results along with robustness checks with different size bin cut-offs are available upon request.

[18] The contrast is less stark in Figure 2.5, where firm size distributions are presented. For the most polluting industry—the Paper and Paper Product Industry—the difference is still evident with around 40% of the U.S firms with more than 400 employees, but less than 10% of those in China. But the pattern is less pronounced for other industries and all polluting industries combined.

the ratios of the concentration indicators of U.S over China for all five top polluting industries are higher than the overall average.

These findings indicate that compared to the U.S, a much larger portion of production is done by small firms in China. Hence the underlying industry structure difference could be a candidate for explaining the high industrial pollution emissions in China.

2.2.5 Size Distribution and Aggregate Pollution

To gain an understanding of the size of the quantitative effect of employment distribution on aggregate pollution, in this section we conduct an accounting exercise. In this exercise, for each polluting industry in China, we fix the level of total industrial production, but replace the employment distribution (which we use as a proxy for production share distribution) with that from the U.S and calculate the subsequent implied level of aggregate industrial pollution. This simple exercise is complicated by the fact that NGSPS only reports the firm-level total value of production but not the number of employees. We use CNEC to overcome this issue. There are many ways to construct the employment-production relationship using CNEC and each method has its own advantages and disadvantages. In this section, we report results where the employment-production relationship is constructed by linear regression. We provide the details of the calculation and results of two alternative estimation methods in Appendix A.

The results are shown in Table 2.5. The numbers reported are the ratio of the aggregate pollution level produced with the U.S employment share distribution over that with the original Chinese distribution.

The results imply that by changing the employment share distribution to that of the U.S, while keeping production at the same level, the aggregate discharge in the

	Paper	Agricultural Food	Textile	Chemistry	Beverage	Average	Reduction
Average Intensity	43.5%	61.1%	97.5%	101.2%	89.0%	67.0%	25.5%

† Note: Please see notes of Table 2.2 for acronyms of industries. For individual industries, the numbers reported are the aggregate pollution from the artificial U.S production structure as percentage from that of China. Column 6 (Ave) calculates the weighted average of these ratios using the percentage contribution in row one of Table 2.2 as weights. Column 7 (Reduc) reports the aggregate reduction, which is simply the average without normalization.

Table 2.5: Size Distribution on Pollution

Paper and Paper Product industry reduces to 43.5% of the original level. The size of such reductions decreases as the overall contributions (i.e., row one in Table 2.2) decrease. The scale of the reduction for each industry is broadly proportional to the industry's contribution to the aggregate pollution. On average, for the top-5 polluting industries, the effect of change in size distribution is reduction of discharge to 67% of the original level. Changing the size distributions of the five industries together will achieve a reduction in total discharge by about 25.5%. Although the exercise here is a crude approximation, it nevertheless shows that size distribution could have a significant impact on the level of aggregate industrial pollution.

2.3 The Model

The accounting exercise in the last section has several limitations. First, the aggregate output is fixed. It is possible that when the size distribution changes, although the pollution intensity decreases, but because of a larger increase in the aggregate output, the aggregate pollution increases as a result. We would like to allow for such a scenario in our analysis. Second, the employment distribution is mechanically changed to that in the U.S. From the accounting exercise alone, we do not know what are the factors that drive the difference between the employment distribution of China and the one of the U.S, nor do we know that by changing these factors, whether the implied employment distribution will in fact become that of the U.S. Third, the firm size and emission intensity relationship is taken as exogenous and invariant. It is possible that changes in the factors that affect the employment distribution also affects the technology choice decisions of the firms, which makes the size-intensity relationship endogenous.

Therefore, to better evaluate the environmental consequences of distortions to firm size, we need a model (i) in which some economic factors affect both aggregate output and pollution; (ii) which reveals what are the factors that affect firm size and how; and (iii) which provides explanation to the size-intensity relationship.

For this purpose, we consider a one sector neoclassical growth model with heterogeneous production units featuring product market frictions, imperfect environmental monitoring and endogenous treatment technology choice. There is a stand-in representative household with a continuum of members in the economy. Each period household members make occupational choices on whether to work as a wage-earner or to become an entrepreneur based on their comparative advantages. We assume there are two types of treatment technologies—dirty and clean. An entrepreneur has to choose between the two upon starting business. The two technologies could be interpreted as the physical and biological technology we discussed in Section II.C. We use physical/dirty and biological/clean interchangeably in the rest of the paper. The environmental regulator imperfectly monitors the installation of clean technology which requires fixed installation costs. If a firm is inspected and is found to be using dirty technology, it receives a penalty.

2.3.1 Setup

Household.—There is a representative household with a continuum of members. Each household member is endowed with z units of managerial talent, $z \sim G(z)$ with support $Z \triangleq [0, \overline{z}]$, where $G(z)$ is the cumulative density and $g(z)$ is the probability density. We assume the support and distribution of z are exogenous. Further we assume that z is fixed once drawn. Household members face an occupational choice decision between worker and entrepreneur. A worker supplies one unit of labor inelastically in exchange for wage income and an entrepreneur rents capital and labor to run a neoclassical firm and earns profits. Let the final product be the numeraire, and R and W be the capital and labor rental price respectively. Firms and capital are owned by the household.

Firms.—Firms combine managerial talent z, capital k and labor n to produce output y according to technology

$$y = F(z, k, n) = z^{1-\gamma}(k^{\alpha}n^{1-\alpha})^{\gamma},$$

where $\gamma < 1$ is the span-of-control parameter. The assumption of decreasing returns to scale with respect to k and n supports a non-degenerate distribution of firms. [19]

The production process generates pollutants e as by-products. The total emission depends on the production scale y and the treatment technology firms use

$$e = E(i, y), \tag{2.2}$$

where $i = 1$ indicates the adoption of clean technology and $i = 0$ otherwise. The installation of the clean equipment incurs fixed cost Rk_E, where we assume that the

[19] We build our model based on Lucas (1978) here. However, all the qualitative properties of our model remain valid if instead we use a model with monopolistic competition [e.g., Melitz (2003)] since the two models are isomorphic [see Appendix I of Hsieh and Klenow (2009)]. In the Melitz model, the decreasing returns to scale come from the concavity in the utility function.

equipment is also rented from the market, just as the production capital k. The benefit associated with the equipment is that firms will not be subject to potential penalties from the regulating agency for using dated treatment technologies. [20]

Regulators.—We assume that the environmental authority monitors the adoption of clean technology by firms with probability p. When a firm that uses dirty production technology gets caught, we assume that a fraction ξ of its total profits is confiscated by the regulating agency. The confiscated profits are distributed to the household as lump-sum transfers so they do not affect the decision problem of household members. This reduced-form way of modeling monitoring policy could for instance be rationalized by a mixed strategy Nash Equilibrium of a behind-the-scenes "monitoring game."

The current industrial pollution management and control framework in China consists of economic incentives and command-and-control instruments. [21] The pollution levy system is the most widely used economic instrument in China. However, it has been widely documented that it places very limited constraints on the pollution emission of the firms because the penalty imposed is very low. Firms only have to pay for the pollutant discharges that go beyond the national standard. The pollutant discharges are self-reported and the truthfulness of the reported discharges is imperfectly examined by the regulators. Further, for firms that discharge multiple pollutants and more than one of the pollutants discharged are above the national standards, firms only have to pay for the one that leads to the highest penalty. We calculate from the CNEC the pollution fees levied on firms as a fraction of total value of output and labor compensation. We find that for firms with strictly positive emission fees, these

[20]We choose to model the installation costs as one-time fixed cost as opposed to fixed cost plus operating cost, or size-dependent fixed cost because the latter two are not supported by empirical evidence. We also assume that the fixed cost is only associated with clean technology. See Section A of the Online Appendix for further details.

[21]See Chapter 5 of World Bank (2001) for a detailed description.

fees only account for 0.06% (median) and 0.3% (mean) of the labor compensation.

Therefore in practice, the environmental agencies rely mostly on the command-and-control instruments. To implement the regulation, field inspections are done by the staff of local environmental agencies. At the firm level, field staff typically check the type of treatment equipment firms installed and test emission intensity of major pollutants. Firms that are found at fault during the field inspection are usually suspended from production for an extended period of time until the issues are resolved. In cases of serious pollution accidents, criminal charges are imposed on the owner of the firm. [22] In our model, the fraction ξ of the profits confiscated is used to approximate these situations. Since according to Table 2.4, the treatment technology used by firms is highly correlated with the pollution intensity, we assume that the regulator in our model checks only the treatment technologies. Although the local environmental agencies also monitor the total amount of discharges, these regulations are usually done at more aggregated level, in most cases based on the provincial-level aggregation. They thus are less relevant to the firm-level decision that we study here. [23]

2.3.2 Product Market Frictions

Chinese firms face large frictions on both the product market and factor markets, and these frictions could have sizeable effects on the size and employment distributions

[22] See Dasgupta *et al.* (2001) for a case study of Zhenjiang.

[23] Firm level inspections in the U.S are also targeted mainly on the adopted treatment technologies. For example, as is stated in the 1977 Amendment of the Clean Air Act, each July every county in the U.S will be classified as either an attainment or a non-attainment county according to their overall pollutants emissions level. Firms in non-attainment counties are subject to substantially stronger environmental regulations. For instance, newly established firms in these counties are required to meet the standard of Lowest Achievable Emission Rate (LAER) which demands the installation of the cleanest possible technology supposedly regardless of costs. While on the other hand, in attainment counties only Best Available Control Technology (BACT) which incorporates cost considerations is required. Similarly, existing firms are also subject to stricter regulations on production and end-of-pipe treatment technologies in non-attainment counties than in the attainment counties. See Becker and Henderson (2000) and the references therein for more details.

of firms as well as other macro aggregates. [24] In this section, we provide evidence on the identification of these frictions and subsequently our modeling choice of the prevailing market frictions.

We follow Hsieh and Klenow (2014) by inferring factor and product market frictions from the average factor products. Standard results of Cobb-Douglas production function implies that the ratio of average factor products over marginal factor products is constant. Therefore, the average factor products also provide information regarding the marginal factor products. [25] In particular, if we let τ_{z_i}, τ_{k_i} and τ_{l_i} be respectively the wedges firms face on the product, capital and labor market, the profit maximization problem of firm i is

$$\pi_i = \max_{k_i, l_i} \left\{ (1-\tau_{z_i}) z_i^{1-\gamma} (k_i^{\alpha} l_i^{1-\alpha})^{\gamma} - (1+\tau_{k_i}) R k_i - (1+\tau_{l_i}) W l_i \right\}.$$

Using the first order conditions, the average product of capital ϕ_k, labor ϕ_l and the capital-labor ratio κ could be expressed as

$$\phi_k = \frac{y}{k} = \frac{(1+\tau_{k_i})R}{\alpha\gamma(1-\tau_{z_i})}, \tag{2.3}$$

$$\phi_l = \frac{y}{l} = \frac{(1+\tau_{l_i})W}{(1-\alpha)\gamma(1-\tau_{z_i})}, \tag{2.4}$$

$$\kappa = \frac{k}{l} = \frac{\alpha}{1-\alpha} \cdot \frac{(1+\tau_{l_i})W}{(1+\tau_{k_i})R}. \tag{2.5}$$

The above equations show that in absence of any market friction ($\tau_z = \tau_k = \tau_l = 0$), ϕ_k, ϕ_l and κ should be equalized across all firms. Equations (2.3) and (2.4) say that firms that face higher capital (labor) and/or product market frictions will demonstrate higher average product of capital (labor). In addition, according to equation (2.5), the

[24]For studies of frictions that Chinese firms face, see Hsieh and Klenow (2009), Brandt *et al.* (2013), Song and Wu (2015), and Tombe and Zhu (2015). Recent studies that focus on the effect of size and employment distributions on macro aggregates include Guner *et al.* (2008), Restuccia and Rogerson (2008), and Adamopoulos and Restuccia (2014) among others.

[25]In particular, using the notations later in this section, we can verify that $\partial y/\partial k = \alpha\gamma z^{1-\gamma} k^{\alpha\gamma-1} l^{(1-\alpha)\gamma} = \alpha\gamma(y/k)$. Similarly, we can show that $\partial y/\partial k = (1-\alpha)\gamma(y/l)$.

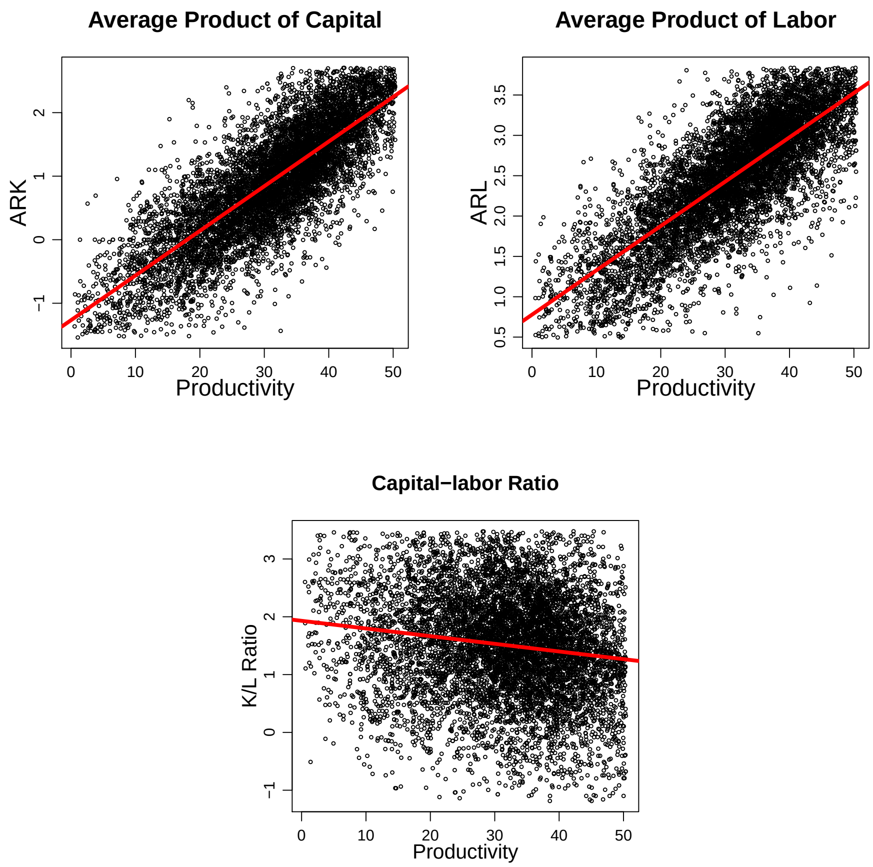

Figure 2.6: Factor and Product Market Frictions

Source: China National Economic Census. All panels are plot in log scale. Lines are least square fit.

capital-labor ratio increases with the relative size of labor to capital market wedge. Using firm-level data on total value of production, book value of capital stock and labor compensation from the CNEC, we calculate z, ϕ_k, ϕ_l and κ for each firm in our sample. Figure 2.6 shows in log scale the scatter-plots of ϕ_k, ϕ_l and κ against firm-level productivity z for the Paper industry.

Two patterns can be observed from Figure 2.6. First, from the two upper panels, we see that both ϕ_k and ϕ_l are positively correlated with z, which shows that more productive firms have higher average product of both capital and labor. Expressed in wedges, this means both $(1+\tau_k)/(1-\tau_z)$ and $(1+\tau_l)/(1-\tau_z)$ are higher for more productive firms. It could be because more productive firms are subject to higher capital or product market frictions or both. Second, from the lower panel, we see that the capital-labor ratio is at best weakly negatively correlated with z. The least squares estimate of the elasticity is -0.0057 and the R^2 is only 0.053. This indicates that the relative wedge firms face on the capital and labor markets do not depend strongly on the idiosyncratic productivity of firms, which in the context of our model implies $1+\tau_k \approx 1+\tau_l$. Since we cannot separately identify the three wedges, for simplicity, we assume $\tau_k = \tau_l = 0$ and attribute all the changes in the average product of factors to wedges in the product market τ_z. [26] Different assumptions on the distribution of frictions across the three markets will not affect our results, but the interpretations need to be changed accordingly. [27]

In the spirit of Restuccia and Rogerson (2008) and Adamopoulos and Restuccia (2014), we implement these idiosyncratic wedges in the model by positing a generic "tax" function that specifies the wedges as a function of firm's productivity z:

$$\tau_z = \max\left\{0, 1-\phi_0 z^{\phi_1}\right\}. \tag{2.6}$$

We assume the taxes collected are returned to household as lump-sum transfers. Anticipating the benchmark calibration in the next section, the wedge function specified in equation (2.6) is increasing and concave in z, with the lower and upper bounds being 0 and 1 respectively. The shape of the function captures the size-dependency

[26]For example, we cannot distinguish between the data generating process we use here and another process where τ_k and τ_l increase simultaneously while τ_y is equal to zero.

[27]We cannot rule out the possibility that the results are driven by measurement error. However, we argue that this does not seem to be the case here. In particular, if y is measured with extreme measurement error, the regression coefficient of ϕ_k over z will be $1-\gamma$. Similarly, if instead k is measured with extreme measurement error, the regression coefficient will be $(1-\gamma)/\gamma$. We calculate ϕ_k and z using different values of γ and the regression coefficients do not vary as predicted by either case.

of the product market frictions where the wedges are higher for larger firms.

There is one difference between (2.6) and the tax function used by Adamopoulos and Restuccia (2014). To model the size dependency, in their specification, the authors use an exponential function as opposed to the power function here. We choose the power function because it is consistent with the log-linearity of ϕ_k (ϕ_l) and z while the exponential function implies a tax scheme that increases much sharper with respect to productivity than the empirical counterpart.

The idiosyncratic τ_z is meant to capture a variety of policies and institutions affecting the profits and, subsequently, the size of the firm. For example, it could be that more productive firms face transportation costs [Adamopoulos (2011)] or local protectionism and trade barriers that impede the inter-regional flow of goods [Young (2000)] when attempting to deliver their products to wider range of areas. It could be that smaller firms are subject to preferential tax treatment. For instance, the value added taxes for firms with annual value of industrial output that is less than CNY 1 million is 3% while firms with production scale larger than CNY 1 million are subject to a 13% tax rate. It is also consistent with a large amount of anecdotal evidence where Chinese entrepreneurs complain about higher administrative costs associated with increasing production scale. [28] Finally, in the language of the trade community, τ_z could also be interpreted as different markups. Melitz and Ottaviano (2008) show that a linear demand yields a demand elasticity that is decreasing in firm size, which subsequently translates into markups that increase with firm size. [29]

In summary, the purpose of τ_z is to capture the empirical regularities in Figure

[28] For example, data from the World Bank Investment Climate Survey which surveys a representative sample (12,000 firms) of China's manufacturing and service firms, report that 12% of the survey respondents named "Anti-competition behaviors by local governments and other enterprises" as the factor that is most damaging to the operation and growth of their firms.

[29] For the markup interpretation, we have to write a monopolistic competition version of our model [Melitz (2003)], instead of the Lucas model we use here.

2.6 in a parsimonious way. We do not intend to evaluate the role of any particular observable product or factor market frictions. In this sense, methodologically we are following the indirect approach as opposed to the direct approach [Restuccia and Rogerson (2013)].

2.3.3 Firm's Problem

Entrepreneurs first decide on which type of treatment technology to use and then on how much to produce. The business profits of a type-z entrepreneur $\pi(z)$ is the maximum over the profits of producing using dirty technology $\pi_0(z)$ and those of using clean technology $\pi_1(z)$:

$$\pi(z) = \max_{i \in \{0,1\}} \{\pi_0(z), \pi_1(z)\}, \tag{2.7}$$

where $i(z)$ is the treatment equipment choice decision.

Firms using clean technology are not subject to environmental penalties, hence their profits are just revenues less costs:

$$\pi_1(z) = \max_{k,n} \left\{(1-\tau_z) z^{1-\gamma} (k^\alpha n^{1-\alpha})^\gamma - Wn - R(k+k_E)\right\}. \tag{2.8}$$

Notice that here the treatment equipment k_E cannot be used to produce the final product. This is consistent with the empirical finding by Shadbegian and Gray (2005).

On the other hand, firms using dirty technology will be inspected by the environmental authority with probability p. Under such circumstances, a fraction ξ of their annual profits will be confiscated. Hence, the profit function is

$$\pi_0^C(z) = (1-\xi)\left[(1-\tau_z) z^{1-\gamma} (k^\alpha n^{1-\alpha})^\gamma - Wn - Rk\right],$$

where the superscript C indicates "caught." While if the firm succeeds in evading the inspection, the profit function is

$$\pi_0^E(z) = (1-\tau_z) z^{1-\gamma} (k^\alpha n^{1-\alpha})^\gamma - Wn - Rk,$$

where the superscript E indicates "evaded." Because we assume perfect risk sharing within the household, these entrepreneurs will not have precautionary motives and will simply maximize the expected profits over π_0^C and π_0^E:

$$\pi_0(z) = \max_{k,n} \left\{ (1-p)\pi_0^E(z) + p\pi_0^C(z) \right\}.$$

Straightforward algebra yields

$$\pi_0(z) = \max_{k,n} \left\{ (1-p\xi) \left[(1-\tau_z) z^{1-\gamma} (k^\alpha n^{1-\alpha})^\gamma - Wn - Rk \right] \right\}. \tag{2.9}$$

2.3.4 Product Market Frictions and Technology Adoption

To clarify the basic mechanics of the model, in this section we analyze firms' optimization problem in a partial equilibrium, where R and W are fixed as given. We prove two results in this section. First, we show that there exists a threshold $\tilde{z}$ such that firms with $z > \tilde{z}$ adopt the clean technology while firms with $z \leq \tilde{z}$ do not. Second, if we denote the previous threshold in environments with and without product market frictions to be respectively $\tilde{z}_f$ and $\tilde{z}_n$, we show that $\tilde{z}_f > \tilde{z}_n$. The first result says that there are returns to scale embedded with clean treatment technologies that are only exploited when firms are large enough. The second result says that by introducing product market frictions, a positive measure of firms that adopt clean technology when there are no frictions do not have the profit margin to benefit from the clean technology, and hence choose to enter the market with dirty technology. Throughout, we assume $0 < \alpha < 1$, $0 < \gamma < 1$, $\phi_0 = 1$ and $1 - \gamma + \phi_1 > 0$. We impose the last inequality because the product market tax specified in (2.6) is imposed on talent z. In order for the benefits of higher talent z (the elasticity of profits to talents $1-\gamma$) to always out-weight the costs (the elasticity of taxes to talents ϕ_1), $1-\gamma+\phi_1 > 0$ must be satisfied. This also rules out the case where the most talented individuals choose to become workers. All proofs are left for the appendix.

Lemma 1 characterizes firms' profit functions in the absence of product market frictions.

Lemma 1. *In an economy with no product market frictions, $\pi_0(z)$ and $\pi_1(z)$ are both increasing and linear with respect to z. In addition, the slope of $\pi_1(z)$ is steeper than that of $\pi_0(z)$:*

$$\frac{\partial \pi_0(z)}{\partial z} = (1-p\xi)\frac{\partial \pi_1(z)}{\partial z}, \qquad \forall z \in Z. \tag{2.10}$$

Lemma 1 highlights the core trade-off of adopting clean technology in our model. Although the up-front fixed costs shift the overall profit function down by Rk_E, firms with clean technology are not subject to the ξ proportion of profits being confiscated. With constant elasticity between capital and labor, the optimizing capital to labor ratio is constant in absence of factor market frictions, therefore entrepreneurs reap economic rents from managerial talents z. These economic rents increase linearly in z, because we assume a constant returns to scale production function.

Since the tax in (2.6) is size-dependent in the sense that more talented entrepreneurs are subject to higher frictions, it is straightforward to show that in an economy with product market frictions, both $\pi_0(z)$ and $\pi_1(z)$ are concave. $\pi_0(z)$ and $\pi_1(z)$ will remain linear if the frictions are uniformly imposed.

Corollary 1. *Suppose the product market frictions are specified as* $\max\left\{0, 1-z^{\phi_1}\right\}$ *with $1-\gamma+\phi_1>0$, then $\pi_0(z)$ and $\pi_1(z)$ are both increasing and concave with respect to z. In addition, the slope of $\pi_1(z)$ is steeper than that of $\pi_0(z)$:*

$$\frac{\partial \pi_0(z)}{\partial z} = (1-p\xi)\frac{\partial \pi_1(z)}{\partial z}, \qquad \forall z \in Z.$$

Since the wage income associated with being a worker is fixed at W, the monotonicity of the profit functions implies that there is a threshold $\hat{z}$ for which all household members with talents higher than $\hat{z}$ choose to become entrepreneurs. Put differently, household members choose their occupations according to their comparative advantages. This is the standard result from the Lucas model. We summarize it below in Proposition 1.

Proposition 1. *There exists a unique threshold $\hat{z}$ such that all household members with $z < \hat{z}$ choose to be workers and those with $z \geq \hat{z}$ become entrepreneurs. Further, $\hat{z}$ is the solution of $W = \pi(\hat{z})$.*

Finally, Proposition 2 summarizes the main result of this section: larger firms adopt more advanced treatment technology and the existence of product market frictions impedes the technology upgrade.

Proposition 2. *Given k_E, W, τ_z and R, there exist unique thresholds $\tilde{z}_n$ and $\tilde{z}_f$ such that:*

(i) In the economy with no product market frictions, entrepreneurs with $z \leq \tilde{z}_n$ produce using dirty technology while those with $z > \tilde{z}_n$ produce using clean technology.

(ii) In the economy with product market frictions, entrepreneurs with $z \leq \tilde{z}_f$ produce using dirty technology while those with $z > \tilde{z}_f$ produce using clean technology.

(iii) $\tilde{z}_n < \tilde{z}_f$, that is, product market frictions impede the technology upgrade.

A graphical illustration of Proposition 2 is shown in Figure 2.7. There are four curves in the figure, $\pi_0^n, \pi_0^f, \pi_1^n$ and π_1^f where superscripts n and f represent whether there are product market frictions and subscripts 0 and 1 represent firms using dirty or clean technology respectively. The elasticity of profits to managerial talents is $1-\gamma$ when clean technology is adopted which is larger than that with dirty technology $1-\gamma+\phi_1$. Therefore, although for less talented entrepreneurs the fixed installation costs of clean technology is not justified, for highly talented entrepreneurs using more advanced

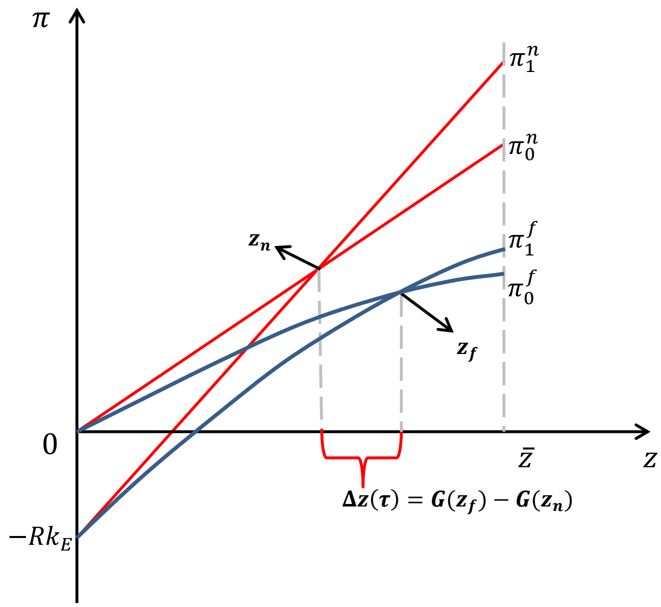

Figure 2.7: The Effect of Product Market Frictions

treatment technologies will eventually pay off. Whether it is worthwhile for firms to pay the installation costs is determined by contrasting the costs with potential profits loss from inspections by the regulator. Since the product market frictions shrink the profit margin by ϕ_1, for some firms their "after-tax" profits do not permit them to exploit the benefits from adopting more advanced technologies, although the "pre-tax" profits do. The ultimate result is that a positive measure $\Delta z(\tau) = G(z_f) - G(z_n)$ of firms that with no frictions would produce using clean technology, in the existence of product market frictions produce using dirty technology.

2.3.5 General Equilibrium

In this section, we specify the household problem and define the general equilibrium to close the model.

Household Optimization.—The household engages in a simple consumption saving problem:

$$\max_{C_t, K_{t+1}} \sum_{t=0}^{\infty} \beta^t U(C_t) \tag{2.11}$$

$s.t.$

$$C_t + K_{t+1} - (1-\delta)K_t = I_t,$$

where C_t is the consumption, K_t is the aggregate capital, β is the discount rate, δ is the depreciation rate, and I_t is household income which we will specify in detail shortly. [30] The solution to (2.11) is the standard intertemporal Euler equation

$$U'(C_t) = \beta U'(C_{t+1})(1 + R_{t+1} - \delta), \tag{2.12}$$

which pins down the equilibrium interest rate.

Household income I_t comes from three sources: wage income, firms' profits and lump-sum transfers associated with product market frictions τ_z and environmental penalties ξ. To characterize I_t, we need some additional notation. We denote $Z_0 = \{z \in [\hat{z}_t, \overline{z}] | \pi_0(z) \geq \pi_1(z)\}$ as the set of firms operating under dirty technology and $Z_1 = \{z \in [\hat{z}_t, \overline{z}] | \pi_0(z) < \pi_1(z)\}$ as the set of firms using clean technology. Notice that for the intermediate case where $0 < \hat{z} < \tilde{z} < \overline{z}$, Proposition 2 implies $Z_0 = [\hat{z}, \tilde{z})$ and $Z_1 = [\tilde{z}, \overline{z}]$. If we let T denote the transfers, which equals taxes in quantity, Equation (2.9) and Proposition 1 then yield:

$$I_t = R_t K_t + W_t G(\hat{z}_t) + \int_{z \in Z_0} \pi_0(z) dG(z) + \int_{z \in Z_1} \pi_1(z) dG(z) + T.$$

where the five terms are capital rental income, wage income and profits from dirty and clean firms, and government transfers. A law of large numbers here guarantees

[30]We assume here that the household does not value environmental quality, but values only consumption. The assumption is innocuous in the competitive equilibrium, since the household has no control over the environmental quality. However, the assumption will matter if a planner's problem in considered.

the ex ante probability of being inspected equals the ex post number of firms that in fact get inspected.

Now we are ready to define the equilibrium. Let Y be the aggregate output and E be the aggregate pollution, the steady state equilibrium of the model is defined as follows.

Definition 1. *A steady state equilibrium in this model is the prices $\{W, R\}$, allocations $\{C, K, Y\}$, firms' policy functions $\{k(z), n(z), y(z), \pi(z)\}$, household's occupational choice $\hat{z}$, firms' technology adoption choice $\tilde{z}$, and aggregate pollutants emission E such that:*

(i) Given factor prices $\{W, R\}$, $\{C, K, \hat{z}\}$ solve the household optimization problem;

(ii) Given factor prices $\{W, R\}$, $\{k(z), n(z), y(z), \pi(z)\}$ and $\tilde{z}$ solve firms' optimization problems;

(iii) Factor prices $\{W, R\}$ clear all markets;

- *Labor Market:*

$$G(\hat{z}) = \int_{\hat{z}}^{\overline{z}} n(z) dG(z).$$

- *Capital Market:*

$$K = \int_{\hat{z}}^{\overline{z}} k(z) dG(z) + k_E \int_{z \in Z_1} dG(z).$$

- *Product Market:*

$$C + K - (1 - \delta) K = \int_{\hat{z}}^{\overline{z}} y(z) dG(z).$$

(iv) The aggregate pollutants discharges are

$$E = \int_{z \in Z_0} e\left(0, y(z)\right) dG(z) + \int_{z \in Z_1} e\left(1, y(z)\right) dG(z).$$

2.4 Calibration

We calibrate our model to the Chinese data. The model period is set to be one year.

Calibration.—Motivated by the empirical evidence in Section I, we assume that the functional form of the pollution intensity function of firms with treatment technology i and production level y is log-linear:

$$\log \frac{e}{y} = \psi_0^{(i)} + \psi_1^{(i)} \log y. \tag{2.13}$$

This specification implies that conditional on treatment technology adopted, there is still "within group" intensity reduction as production scale increases. Since our model includes only the choices of firms on the end-of-pipe treatment technologies, in the quantitative exercises, we capture the decrease in pollution intensity in a reduced-form way. Equation (2.13) implies that the actual discharge is

$$e = E(i, y) = \mathrm{e}^{\psi_0^{(i)}} y^{1+\psi_1^{(i)}}. \tag{2.14}$$

Our model features only two broad categories of treatment technologies as opposed to five in the data. We interpret the clean technology in our model as biological technology and dirty technology as the remaining types. Since the aggregated installation costs of the physical equipment are less than 9% of those of the biological equipment, we assume only the installation of biological technology incurs the fixed cost Rk_E.

Because the firm size distribution in the model is affected by both the talent distribution $G(\cdot)$ and the product market frictions τ_z, our identification assumption is such that parameters governing τ_z [ϕ_0 and ϕ_1 in Equation (2.6)] are calibrated according to the empirical regularities in Section II.B (explained in detail shortly after) and given τ_z, $G(\cdot)$ is set to match the firm size and employment distribution in China. We choose the pooled polluting industries as our calibration targets. The employment and firm size distributions of these industries pooled together are shown in the lower-right panels of Figures 2.4 and 2.5. The firm size distribution resembles a log-normal distribution, but also demonstrates a considerable degree of employment concentration at very large firms. It is well documented in the literature that the commonly used log-normal distribution does a reasonably good job at matching the distribution of the bulk of small and medium-sized firms, but does not support the concentration of employment. The heavy right tail is crucial to our evaluation because these are the firms that are producing with clean technology. Since τ_z is levied based on the productivity z, we assume the distribution of $z' = (1-\tau_z)z^{1-\gamma}$ to be a combination of two components. The first is a log-normal distribution with mean μ, standard deviation σ and total probability mass $1 - g_{max}$ that accounts for the bulk of small and medium firms. The second is an atomic with value z'_{max} and measure g_{max} which accounts for the very large firms. [31] The talent z is then calculated by

$$z = \left(z' \phi_0^{1/(\gamma-1)}\right)^{\frac{1-\gamma}{1-\gamma+\phi_1}},$$

which gives us $G(z)$.

Therefore, we are left with total of 17 parameters to calibrate: discount factor β, production technology parameters $\{A, \delta, \alpha, \gamma\}$, treatment technology parameters $\{\psi_0^{(0)}, \psi_1^{(0)}, \psi_0^{(1)}, \psi_1^{(1)}, k_E, p\xi\}$, product market frictions $\{\phi_0, \phi_1\}$ and distributional parameters $\{\mu, \sigma, z'_{max}, g_{max}\}$. The general strategy of our calibration involves assigning values to some parameters based on a priori information in the data and calibrate the rest jointly such that the distance between the moments from the model and the data is minimized.

Eight of the seventeen parameters can be determined exogenously. We set the depreciation rate δ to 10% [Song *et al.* (2011)]. To get estimates of $\{\psi_0^{(0)}, \psi_1^{(0)}, \psi_0^{(1)}, \psi_1^{(1)}\}$,

[31] This strategy follows Guner *et al.* (2008) and is quite popular among macroeconomic studies on wealth distribution, see for example **?**.

we repeat the exercises in Section I.B for firms using physical and biological equipment separately. The estimates are $\psi_0^{(0)} = -3.5795$, $\psi_1^{(0)} = -0.4149$, $\psi_0^{(1)} = -4.4270$ and $\psi_1^{(1)} = -0.3410$. In the context of our model, these estimates suggest that on average, for two firms with the same level of production but different treatment technology, the firm that uses biological technology discharges 40% to 60% less pollutants than the firm equipped with physical technology. We use information on the average products of capital and labor to calibrate the tax function. Equation (2.3) suggests that the elasticity of ϕ_k to $1-\tau_z$ is equal to unity. Therefore ϕ_1 is equal to the elasticity of ϕ_k to z. We therefore calculate ϕ_k and z according to Section II.B, with $R = 0.1$ [Hsieh and Klenow (2009)] and the same γ used later when we are calibrating the model to match the firm size and employment distributions. We then run a regression

$$\log \phi_{k_i} = \kappa_0 + \phi_1 \log z_i + \varepsilon_{z_i},$$

which gives us the value for $\phi_1 = -0.03$. Given ϕ_1, we then calibrate ϕ_0 such that the average tax burden in the economy equals the value added tax imposed on Chinese manufacturing firms in the data, which is 13%. This gives us the value of $\phi_1 = 1.15$. We choose to target only the value added tax rate and not include other frictions because as is shown in Proposition 2, the gains in output (and capital and consumption as well) from eliminating the product market friction are increasing in the average level of frictions. Targeting a higher average "tax" rate increases such gains monotonically, and will not affect the results qualitatively. We therefore choose a conservative target for the average level of frictions. We set $A = 1$ as normalization.

The remaining parameters are calibrated jointly. The calibration involves two layers: an outer layer loops over the parameterization of $G(z)$ and an inner layer solves the model given $G(z)$. In the inner layer, first we approximate $G(z)$ with 5,000 grid points. We then choose β and α to match respectively the capital-output ratio of

1.65 and capital share of 0.5 [both are taken from Bai *et al.* (2006)] in China. We set $p\xi$ and k_E such that the total treatment equipment investment is equal to 1% of the total output, and the fraction of firms adopting biological equipment equals the empirically observed level of 57%. The value of γ is set such that the difference between the numbers of firms fall in each bin of the employment and firm size distributions generated by the model, and those in the data is minimized. More specifically, if we let s_q and $\hat{s}_q$ be the number of firms in each bin q (in total ten of them) calculated from the data and from the model respectively, γ^* solves

$$\gamma^* = \underset{\gamma}{\operatorname{argmin}} \sum_{q=1}^{10} \left(s_q - \hat{s}_q\right)^2 . \tag{2.15}$$

Notice that for each combination of $\{\mu, \sigma, z_{max}, g_{max}\}$, there is one corresponding γ^*. Therefore, in the outer layer, we use a multi-dimensional search process to choose the combination of $\{\mu, \sigma, z_{max}, g_{max}\}$ that minimizes the minimum distances from the inner layer. Put differently, the outer layer picks the minimum of the right hand side values of Equation (2.15) when evaluated at γ^*. The model parameters along with their targets and calibrated values are listed in Table 2.6.

Discussion.—The calibrated model matches very well the capital share, capital-output ratio, the fraction of treatment equipment in total output and the fraction of firms adopting biological equipment. The calibrated value of returns to scale γ lies within the empirically estimated range. [32] The calibrated value of $p\xi$ means that in expectation, the penalty to firms using less advanced treatment technology equals 20.5% of their annual output value.

Figure 2.8 shows graphically the firm size (left panel) and employment share dis-

[32] The values previously used in the macro literature range from 0.85 [Atkeson and Kehoe (2005)] at the lower end to 0.95 [Bartelsman *et al.* (2013)] at the upper end. Estimations from micro-level data yield similar results, for example Olley and Pakes (1996) estimated the value to be between 0.8 to 0.9 for the U.S telecommunications equipment industry, depending on the particular econometric specifications.

Parameter		Value	Targets
Production	A	1	Normalization
	δ	0.1000	Depreciation Rate
	α	0.5376	Capital Share 0.5
	γ	0.9300	Size Distribution†
Treatment	$\psi_0^{(0)}$	-3.5795	Physical Intensity-output Elasticity
	$\psi_1^{(0)}$	-0.4149	
	$\psi_0^{(1)}$	-4.4270	Biological Intensity-output Elasticity
	$\psi_1^{(1)}$	-0.3410	
	k_E	4.1500	Envir.capital-output ratio 1%
	$p\xi$	0.205	Frac.Firms Use Bio 57%
Frictions	ϕ_0	1.15	Average Value Added Tax 13%
	ϕ_1	-0.03	Avg.Factor.Prod-Prod Elasticity
Preference	β	0.8750	Capital-output Ratio 1.65
Talents	μ	-2.4567	Size Distribution†
	σ	4.0020	
	z_{max}	10820.4	
	g_{max}	0.00048	

† Note: Jointly calibrated.

Table 2.6: Parameterization

tributions (right panel) in the model and in the data. While not perfect, overall the model does a reasonable job in matching the two distributions given that there are five degrees of freedom. The mean (59.27) and median (22.95) of the firm size distribution, which are not directly targeted in the calibration, also match well with their empirical counterparts, where the mean and median are equal to 59.05 and 20 respectively. The challenges of calibrating the model to simultaneously match the employment and size distributions are as follows. First, in order to create the large

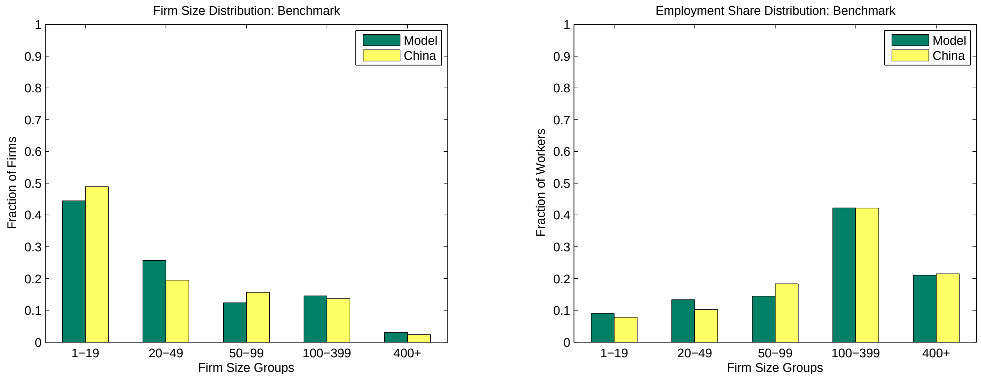

Figure 2.8: Model Fit: Benchmark

firms in the model, we need not only very talented entrepreneurs who are willing to hire a lot of employees, but also the wage these entrepreneurs face has to be kept at a low level to make them able to actually hire the desired amount of workers. Therefore, the average talent in the economy must be low. If we have a wide range of employment level to cover, the properties of the distributions at the right tail of the log-normal are difficult to control. It is for this particular reason that we introduce an atom to the distribution. Second, since for the five industries we are studying in this paper the employment concentration level is very high, to generate comparable level of concentration, the returns to scale (γ) must also be high given that the average talent is low as we just discussed. This will significantly jeopardize the firm size distribution as the big firms enjoy too many technological advantages. The fit of our model is thus the compromise of these two forces. [33]

The identification assumption of our benchmark calibration is that given taxes τ_z, the talent distribution is identified by the size distribution in the data. Another

[33]Models where these two forces are not very antagonistic to each other usually have much better fit. For instance, Guner *et al.* (2008) study the whole U.S business sector which has narrower employment span and lighter employment concentration. In particular, the largest group they are targeting is firms with more than 100 employees. On the contrary, the largest group we are targeting is 400+ employees, which is significantly larger. In another paper, Adamopoulos and Restuccia (2014) assume away the selection mechanism in the model which, put differently, mutes the general equilibrium feedback through wages. This relaxes the restriction on the average talents considerably.

commonly used calibration strategy in the literature takes the U.S as an undistorted economy (which means $\tau_z = 0$ in our model) and calibrates $G(z)$ by matching the size distribution from the tax-free economy to the U.S data. Taxes are then introduced such that certain moments of the size distribution of the country in interest are matched. [34] If we calibrate our model in this way, the underlying assumption would be that the talent distribution of the entrepreneurs in China is the same as that in the U.S and all the differences between the size distributions of China and the U.S result from taxes. [35] From our perspective, both strategies have their strength and weakness, and compromise to data limitations in different ways. The quantitative results of our paper are to largely driven by size and shape of τ_z. Therefore, as long as alternative calibrations imply tax schemes that are similar to what we use in our benchmark calibration, the quantitative aspects of our results will hold as well. [36] We choose to calibrate our model to the Chinese economy directly because better empirical evidence (Section II.B) is available to us.

2.5 Quantitative Results

We use the calibrated model as a framework for understanding the effects of firm size distribution on industrial pollution. We conduct two experiments. In experiment (i), we eliminate all the product market frictions by setting $\tau_z = 0$. The experiment could be interpreted as reductions in inter-regional trade barriers, improvements in transportation infrastructure, decreases in tax burdens, etc. Since we are following the *indirect* approach, we do not have empirical evidence of how much improvements on observable frictions (for instance the transportation infrastructure) reduces τ_z by

[34] See for example, Guner *et al.* (2008), Hsieh and Klenow (2009, 2014), and Adamopoulos and Restuccia (2014) among others.

[35] The evidence regarding this point is mixed, see Figure 2 in Bloom and Van Reenen (2010).

[36] In an earlier version of the paper where endogenous treatment technology choice is not explicitly modeled, we find that this is indeed the case.

how much. The effects of the policy are then assessed by the changes in the average firm size across steady states as is done in Guner *et al.* (2008). In experiment (ii), we increase the monitoring $p\xi$ such that the fraction of firms using biological equipment reaches the same level as in experiment (i). With this experiment, we approximate the current environmental policy that punishes firms using less advanced treatment technology. We contrast results from the two experiments to illustrate the different effects of these two types of policies.

2.5.1 *Less Frictions vs. Stronger Regulation*

In this section, we compare the effects of the two types of policies, namely reducing frictions and intensifying regulation. We start by describing the results of the two experiments. Table 2.7 contains results that characterize the steady states. Columns labeled (i) and (ii) refer to experiments (i) and (ii) respectively.

Elimination of Frictions.—The core mechanism that generates the results of experiment (i) is the change of size distribution that is driven by the general equilibrium wage effect. Since the tax scheme τ_z is assumed to be size-dependent, which imposes larger frictions on more productive firms, in the benchmark economy the market share of these highly productive firms is severely restrained. The elimination of τ_z removes these constraints. As a result, the previously suppressed factor demand increases considerably. In column (i) of Table 2.7, this shows up as a 61% increase in aggregate capital and a 28% increase in wage (output per worker). The size-dependency of τ_z also implies that the situation of small unproductive firms improves to a lesser extent than that of the large productive ones. Many small unproductive firms that previously survived because of the low prevailing wage now lose their profit margins. The owners of these firms therefore find it more profitable to work for the more productive firms. This selection mechanism explains the 125% increase in average

Statistics	Benchmark	(i)	(ii)
Aggregate Output	100.00	129.72	100.12
Capital	100.00	161.25	100.14
Consumption	100.00	123.45	100.11
Output per Worker	100.00	128.49	99.97
Output per Firm	100.00	298.78	109.76
Average Talent	100.00	224.67	109.49
TFP	100.00	102.00	100.14
Number of Firms	100.00	43.42	91.22
Mean Size	59.27	137.81	65.07
Median Size	22.95	41.83	26.72
Aggregate Pollution	100.00	78.95	90.75
Average Intensity	100.00	60.86	90.65
Biological Share	57.40	85.29	85.15
Monitoring	20.50	20.50	32.50

† Note: All values reported are in percentage points except mean and median size, which are numbers of workers in absolute term.

Table 2.7: Aggregate and Productivity Effects

talent of active entrepreneurs, the 57% decrease in the number of active firms, the three-fold increase in output per firm and the increase in the mean and median size of the firms. Beyond the increase in the cutoff we just described, production is also more concentrated at firms with high productivity among the remaining firms. We define the *extensive* margin as the selection of active entrepreneurs and the *intensive* margin as the production distribution among the active firms. The first two rows of Table 2.8, which report the output share accounted for by firms with productivity in different quantiles, show this intensive margin. The overall changes of the size and employment distribution can be seen from the left two panels of Figure 2.9. We see

Economy	QU_1	QU_2	QU_3	QU_4	QU_5
Benchmark	2.69	4.19	7.29	16.55	69.28
Case (i)	1.50	2.83	6.34	18.06	71.27
Case (ii)	2.93	4.47	7.77	17.37	67.46

† Note: QU_1 to QU_5 represent respectively the first to the fifth quintile.

Table 2.8: Output Share by Different Managerial Talents Quintile

clearly that firms in the top group expand considerably at the expense of firms in the bottom three groups. Therefore, elimination of τ_z improves resource allocation on both the extensive and intensive margin. Consistent with the findings in Guner *et al.* (2008), size distribution in models with firm selection exerts very limited influence over TFP, which is defined as the Solow residual following standard growth accounting literature [Hall and Jones (1999)]:

$$\text{TFP} = \frac{Y}{(K^{\alpha}N^{1-\alpha})^{1-\gamma}}.$$

Although aggregate output increases by 30%, because the average pollution intensity decreasing more (by 40%), the aggregate pollution in fact decreases by 20%. The decline in average intensity enters in both the production stage and the treatment stage. In both stages, changes in the size distribution assume a key role. Since the production in the no friction economy shows higher degree of concentration in large productive firms, the fact that the two within group elasticities $\psi_1^{(0)}$ and $\psi_1^{(1)}$ are negative implies mechanically a decrease in pollution intensity. The effect of size distribution on a firm's choice of the end-of-pipe treatment technology works exactly as described in Proposition 2. The elimination of τ_z increases significantly the profits of firms, which strengthens the economic incentive of adopting more advanced treatment technology. It is important to bear in mind that both the elimination of τ_z itself and the subsequent decrease of wages contribute to the strengthening of the economic

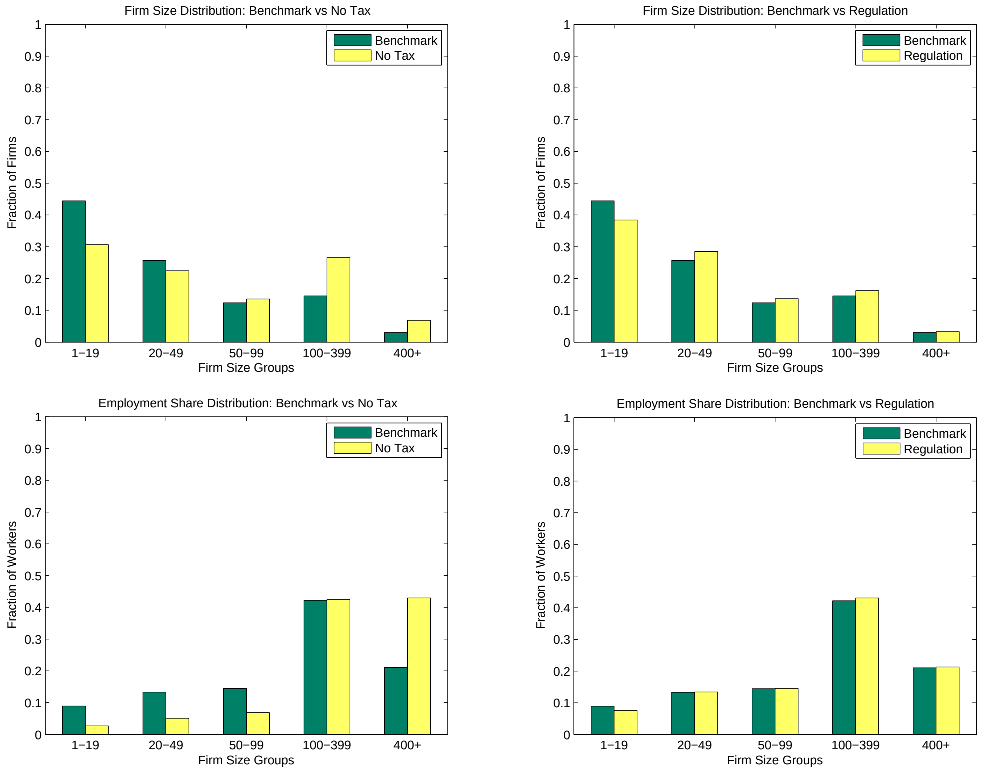

Figure 2.9: Size Distribution: Less Frictions vs. Stronger Regulations

incentive.

To evaluate the relative contribution of reduction in the production stage and in the treatment stage, we assume artificially that $\psi_0^{(0)} = \psi_0^{(1)}$ and $\psi_1^{(0)} = \psi_1^{(1)}$ which means biological equipment has the same technical features as physical equipment. We apply this modification to both the benchmark case and the no tax case. We set the pollution intensity and aggregate pollution in the first case to be the new benchmark. The difference between the above two cases shows the effect from purely changing the size distribution, which is equal to 61% for the intensity and 79% for the aggregate pollution. We interpret these numbers as reduction of industrial pollution in the production stage. We then express the intensity and pollution in case (i) as a percentage of the new benchmark. The numbers are respectively 45% for intensity and 58% for aggregate pollution. Therefore, in the context of our model, about 30% of the decrease in pollution intensity and 50% of the decrease in aggregate pollution are from the treatment stage.

Strengthening of Regulation.—The intensification of regulation affects the decisions of firms directly through technology adoption requirement and indirectly through the general equilibrium wage feedback. Unlike in experiment (i), these two effects do not affect the size and employment distributions in equilibrium by too much, as can be seen from the two right panels in Figure 2.9. There are two reasons behind this. First, large firms that already have adopted the clean technology in the benchmark are not affected by the changes in the policy. Second, the installation costs of the clean technology in our benchmark calibration are small. As a result, they do not divert a significant portion of the firms' resources from productive use, and therefore do not affect the optimal operating scale of firms by much. As a result, in column (ii) of the top panel of Table 2.7, the macro aggregates barely change compared to the benchmark case.

However, despite a tiny decrease (0.03%) in wages, there is selection which to some extent increases the average size and productivity of active firms. The underlying mechanism here is that an increase in regulation decreases the expected returns from being an entrepreneur, which drives out the least productive firms that cannot afford the installation of a clean technology. [37] These household members then choose to become workers, which increases the labor supply and suppresses equilibrium wage rates. The remaining more productive and hence larger firms benefit from

[37] This prediction matches well the policy practice in China. For example, during 2004 to 2008, the emission of major air pollutants together with industrial production have declined significantly. The reduction of per unit GDP emission for these pollutants were 35% for SO_2, 29% for Black Carbon and 31% for CO (cf. Lin *et al.* (2014)). During that period of time, 34 million kW coal-burning electric generating sets were directly shut down, which amounts to 6.18% of the total electric production in 2013 (*National Development and Reform Commission* [2009] Decree 4).

the reduction in wage rates and expand their operating scale, which explains the increases in size and productivity. Graphically this is reflected in Figure 2.9 as a contraction of group "1-19" both in the size and employment distributions and expansion in other four groups.

Firms in the four expanding groups are not affected equally though. To see this, we refer to row (ii) in Table 2.8. As is shown in the experiment (i), the most efficient way of allocating the talents is to shift the production to most productive firms, increasing the share of output accounted for by firms with productivity in the top quantiles. However, row (ii) says the opposite. Comparing with the benchmark case, the proportion of production accounted for by the lower quantiles increases while that of the top quantile decreases. Therefore, although strengthening the monitoring improves the allocation of managerial talents on the extensive margin, the allocation on the intensive margin worsens. Both effects are small here because in our benchmark calibration k_E is small. In Section IV.C, we discuss the case where k_E is set to a higher level.

Since the aggregate output only increases slightly, quantitatively the decreases in aggregate pollution and intensity are almost the same. In this experiment, they decrease by about 10%. Since the size and employment distributions do not change much here, most of the decrease comes from the treatment stage. In fact, if we repeat the decomposition exercise we did for experiment (i), 92% of the reductions in both intensity and aggregate pollution stem from by the adoption of more advanced technologies.

Comparing the Two Policies.—The lesson we learn from the quantitative exercise is that if resources are devoted to smoothing the frictions in the economy instead of being used to intensify regulations, reductions of pollution in both the production and treatment stage arise naturally as equilibrium outcomes. In fact, the effect of

Statistics	Benchmark	(i)	(i')
Aggregate Output	100.00	129.72	105.33
Capital	100.00	161.25	106.38
Consumption	100.00	123.45	105.12
Output per Worker	100.00	128.49	104.34
Output per Firm	100.00	298.78	242.61
Aggregate Pollution	100.00	78.95	70.60
Average Intensity	100.00	60.86	67.02
Biological Share	57.40	85.29	72.21

† Note: All values reported are in percentage points.

Table 2.9: The Effect of Size Distribution

regulation policies such as government campaigns are often ineffective for their ease of rebound. Since the reduction of pollution could be achieved even with less regulations under the case of no frictions, directing resources at improving economic efficiency is more likely to be effective in China. Put differently, because elimination of τ_z increases output and decreases pollution simultaneously, our results actually suggest that economic development and environmental protection are not necessarily in sharp conflict with each other.

2.5.2 The Effect of the Size-Dependency of Distortions

To further isolate the effect of size distribution, we solve a version of the model where the product market frictions are imposed uniformly over all firms in the economy. More specifically, in these exercises, we set values of τ_z such that the total amount of taxes collected is the same as in the benchmark case with size-dependent τ_z. The implied tax rate $\tau_z = 18\%$ is higher than the average tax rate in the size-dependent case, 13%. The results of the experiment are summarized in Table 2.9.

Columns benchmark and (i) are results of the benchmark calibration and those from setting $\tau_z = 0$ respectively. We label the uniform tax case as (i'). By comparing benchmark with (i'), we are able to assess the effect of the size-dependency of τ_z. Similarly, a comparison of (i) with (i') reveals the effect of levying a flat tax. First notice that a uniform τ_z imposed on all firms does not change the extensive margin comparing to the zero τ_z situation, therefore measures of average talents, number of firms, mean/median size of firms and TFP are not affected [see Guner *et al.* (2008) for details].

Aggregate output, capital, consumption, output per worker and output per firm in case (i') all increase comparing to the benchmark case, which is consistent with findings about size-dependent distortions in the literature. What is different here is that since the uniform tax rate needed to generate the same tax revenue is relatively high (the tax rate for the largest firm in the size-dependent case is 25%), the tax itself still results in a considerable amount of output loss. The source of output loss in our model is the misallocation of the entrepreneurial talent z. However, for the average pollution intensity, much of the reduction is achieved through the elimination of the size-dependency of τ_z. The adoption rate of the clean technology increases by 15 percentage points in case (i'), which is 53% of the total increase resulting from the complete elimination of τ_z. 84% of the total decline in aggregate pollution in the zero τ_z case is achieved by simply removing the size-dependent feature of τ_z.

Discussion.—The finding that size-dependency of τ_z affects the economic efficiency moderately, but the average pollution intensity considerably could be well explained by the theory established in Hopenhayn (2014). In particular, he shows that size-dependency of the policy does not necessarily imply large distortions. What matters for the size of the distortion is the total amount of resources that are affected, not these resources belong to which firms. Studies like Restuccia and Rogerson (2008) and Guner *et al.* (2008) find that size-dependent policies affect economic efficiency more than their size-independent counterparts because those size-dependent policies happen to lead to large amount of resources being affected. The fact that in the flat tax case we impose a fairly large τ_z explains why removing size-dependency of τ_z does not improve the efficiency of the economy by much. However, for the sake of industrial pollution, size distribution does matter. Since larger firms produce in a cleaner way, if the majority of the production is done by large as opposed to small firms, pollution will in fact decrease. Since the size distribution of the firm is purely affected by the size-dependency of τ_z and not by the size of the distortions (in the definition of Hopenhayn (2014)), although size-dependency does not necessarily imply significant efficiency loss, it necessarily results in aggravation of pollution.

2.5.3 Environmental Policy and Size Distribution

In our benchmark calibration, we choose a relatively small k_E at about 2.5 times the equilibrium wage of a typical worker. This limits the extent to which environmental policy could affect the real economy. In this section, we show that when k_E is increased to ten times the value used in the benchmark calibration, environmental policy has sizable effect on the allocation of talents. In particular, we show that although environmental policy improves the efficiency through selection at the extensive margin (occupational choice), the allocation at the intensive margin (production distribution among active firms) worsens. To show this, we consider two scenarios. In the first case, we solve the model again with all parameters remaining at the benchmark calibration level, but increase k_E to 41.5. With no changes in the regulation intensity, the adoption rate of clean technology decreases. Therefore, in the second case, we increase $p\xi$ such that the adoption rate in the benchmark case (57%) is restored. We label these two experiments by (iii) and (iv).

Statistics	Benchmark	(iii)	(iv)
Aggregate Output	100.00	100.00	100.41
Capital	100.00	100.28	101.45
Consumption	100.00	99.94	100.21
Output per Worker	100.00	100.00	99.63
Output per Firm	100.00	100.00	187.70
Average Talent	100.00	100.00	184.14
TFP	100.00	100.00	100.81
Number of Firms	100.00	100.00	53.50
Mean Size	59.27	59.27	111.65
Median Size	22.95	22.95	63.24
Aggregate Pollution	100.00	125.64	87.56
Average Intensity	100.00	125.64	87.20
Biological Share	57.40	8.89	57.06
Monitoring	20.50	20.50	73.50

† Note: All values reported are in percentage points except mean and median size, which are numbers of workers in absolute term.

Table 2.10: Aggregate and Productivity Effects: Higher k_E

The results of experiments (iii) and (iv) are shown in Table 2.10. First we notice that the size distribution as well as the efficiency of the economy stay virtually the same as the benchmark case. The clean technology adoption rate falls to 9%, since a large number of firms now find it unprofitable to use more advanced treatment technology and a higher level of pollution follows. This result suggests that subsidies to treatment technology upgrading can work as a substitute for stronger regulation.

If we increase the regulation to the level such that the original technology adoption rate is restored, as opposed to the results of case (ii) where size distributions and the allocation of resources are only mildly affected, the changes in experiment (iv)

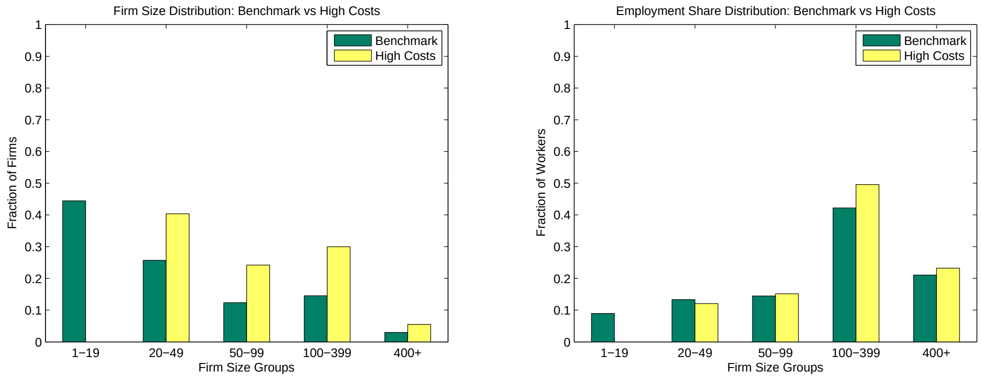

Figure 2.10: Environmental Policy and Size Distribution

Economy	QU_1	QU_2	QU_3	QU_4	QU_5
Benchmark	2.69	4.19	7.29	16.55	69.28
Case (iv)	4.44	6.63	10.98	21.86	56.10
Case (i)	1.50	2.83	6.34	18.06	71.27

† Note: QU_1 to QU_5 represent respectively the first to the fifth quintile.

Table 2.11: Allocation of Production at the Intensive Margin: High k_E

are enormous. The size and employment distributions for case (iv) are shown in Figure 2.10. In contrast to the size and employment distributions in the benchmark case, all firms in the first group are driven out. This is also reflected in column (iv) of Table 2.10 as increases in average talent and mean/median size of the firms. Therefore environmental policy improves resource allocation at the extensive margin by forcing small unproductive firms to quit the market. However, the gains from these improvements are limited because the allocation of resources at the intensive margin worsens. Table 2.11 shows the allocation of resources at the intensive margin. The efficient allocation at the intensive margin is achieved in the no size-dependent τ_z case, which is shown again here in row (i). Instead of an allocation of production toward larger firms, as in the efficient case, here the production of firms in the bottom four quantiles expands significantly at the expense of production share of the most productive firms. As a result, much of the gains in economic efficiency are offset by the worsening of resource allocations at the intensive margin. Although the level of industrial pollution gets lower, the reduction could be much larger if the allocation at the intensive margin was also improved, as is in the case of elimination of τ_z.

2.6 Conclusion

In this paper, using a unique micro-level manufacturing census, we find a strong negative correlation between firm size and pollution intensity in production. We also document substantial differences in firm size and employment distributions between China and the U.S. We find empirical evidence which suggests that size-dependent product market frictions contribute significantly to these observed differences. We use a quantitative framework to organize these empirical regularities, and to study the implication of firm size distribution on industrial pollution at the aggregate level. Quantitative analysis shows that firm size distribution has a sizable impact on industrial pollution. Our results imply that traditional productivity-oriented measures of the costs of size-dependent policies underestimate the true costs of these policies, because industrial pollution, which arguably affects households' welfare, is not accounted for in previous studies. Furthermore, our results suggest an alternative approach to reducing the discharge of industrial pollutants, which focuses on elimination of the economic frictions.

GDP-oriented promotion scheme is often identified as the cause of China's heavy industrial pollution. Our paper shows that growth-enhancing policies that smooth the frictions in the market could foster economic growth and reduce pollutants discharge at the same time. To this end, identifying observable factors that generate the product market frictions and designing optimal environmental policies are very important directions to pursue in future studies. Our model is constructed to facilitate steady state comparisons and, as a result, abstracts from features that could be pertinent for the analysis of short-run policy effects. Extending our model to allow for short-run dynamic path and firm's life-cycle analysis is important to further our understanding of the costs and benefits of certain pollution reduction policies.

In our paper, we focus on industrial water pollution (more specifically Chemical Oxygen Demand) and the case of China. However, the conclusions in our paper could be generalized to other pollutants and cross-country comparisons that involve more countries. For instance, small capacity coal-burning plants are widely acknowledged to be the main source of the emission of sulfur dioxide resulting in acid rain that affects a wide range of areas in China. There are studies using micro-level manufacturing census of firms' production and emissions of other countries [Barrows and Ollivier (2014), Shapiro and Walker (2015) and Dardati (2014)], cross-country comparisons are therefore important directions to pursue. It is also interesting to study the geographical concentration of firms and pollution using a macroeconomic framework. We leave these extensions to future research.

REFERENCES

Acemoglu, D., P. Aghion, L. Bursztyn and D. Hemous, "The environment and directed technical change", American Economic Review **102**, 1, 131–166, URL http://www.aeaweb.org/articles.php?doi=10.1257/aer.102.1.131 (2012).

Adamopoulos, T., "Transportation costs, agricultural productivity and cross-country income differences", International Economic Review **52**, 2, 489–521 (2011).

Adamopoulos, T. and D. Restuccia, "The size distribution of farms and international productivity difference", American Economic Review **104**, 6, 1667–1697 (2014).

Atkeson, A. and A. T. Burstein, "Innovation, firm dynamics, and international trade", Journal of Political Economy **118**, 3, 433–484 (2010).

Atkeson, A. and P. J. Kehoe, "Modeling and measuring organization capital", Journal of Political Economy **113**, 5, 1026–1053 (2005).

Axtell, R. L., "Zipf distribution of U.S. firm sizes", Science **293**, 5536, 1818–1820 (2001).

Bai, C.-E., C.-T. Hsieh and Y. Qian, "The return to capital in china", Brookings Papers on Economic Activity **2006**, 2, 61–88 (2006).

Barrows, G. and H. Ollivier, "Does trade make firms cleaner? Theory and evidence from Indian manufacturing", Manuscript (2014).

Bartelsman, E., J. Haltiwanger and S. Scarpetta, "Cross-country differences in productivity: The role of allocation and selection", American Economic Review **103**, 1, 305–334, URL http://www.aeaweb.org/articles.php?doi=10.1257/aer.103.1.305 (2013).

Becker, R. and V. Henderson, "Effects of air quality regulations on polluting industries", Journal of Political Economy **108**, 2, 379–421, URL http://www.jstor.org/stable/10.1086/262123 (2000).

Benabou, R., "Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?", Econometrica **70**, 2, 481–517, URL https://ideas.repec.org/a/ecm/emetrp/v70y2002i2p481-517.html (2002).

Bento, P. and D. Restuccia, "Misallocation, Establishment Size, and Productivity", , tecipa-517, URL https://ideas.repec.org/p/tor/tecipa/tecipa-517.html (2014).

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie and J. Roberts, "Does management matter? evidence from India", The Quarterly Journal of Economics **128**, 1, 1–51 (2013).

Bloom, N., C. Genakos, R. Martin and R. Sadun, “Modern management: Good for the environment or just hot air?”, Economic Journal **120**, 544, 551–572 (2010).

Bloom, N. and J. Van Reenen, “Why do management practices differ across firms and countries?”, Journal of Economic Perspectives **24**, 1, 203–224 (2010).

Brandt, L., J. V. Biesebroeck and Y. Zhang, “Creative accounting or creative destruction? firm-level productivity growth in chinese manufacturing”, Journal of Development Economics **97**, 2, 339–351, URL `http://www.sciencedirect.com/science/article/pii/S0304387811000216` (2012).

Brandt, L., T. Tombe and X. Zhu, “Factor market distrotions across time, space and sectors in china”, Review of Economic Dynamics **16**, 1, 39–58 (2013).

Clarke, S., “New deal regulation and the revolution in american farm productivity: A case study of the diffusion of the tractor in the corn belt, 1920—1940”, The Journal of Economic History **51**, 1, 101–123, URL `http://www.jstor.org/stable/2123053` (1991).

Copeland, B. R. and M. S. Taylor, “Trade, growth, and the environment”, Journal of Economic Literature **42**, 1, 7–71, URL `http://www.jstor.org/stable/3217036` (2004).

Dardati, E., “Pollution permit systems and firm dynamics: How does the allocation scheme matter?”, International Economic Review (2014).

Dasgupta, S., B. Laplante, N. Mamingi and H. Wang, “Inspections, pollution prices and environmental performance: Evidence from china”, Ecological Economics **36**, 3, 487–498 (2001).

Dasgupta, S., R. E. Lucas and D. Wheeler, “Small plants, pollution and poverty: New evidence from brazil and mexico”, World Bank (1998).

Davis, S. J., J. Haltiwanger, R. Jarmin and J. Miranda, “Volatility and Dispersion in Business Growth Rates: Publicly Traded versus Privately Held Firms”, in “NBER Macroeconomics Annual 2006, Volume 21”, NBER Chapters, pp. 107–180 (National Bureau of Economic Research, Inc, 2007), URL `https://ideas.repec.org/h/nbr/nberch/11178.html`.

Dinlersoz, E., J. Greenwood and H. Hyatt, “Who do Unions Target? Unionization over the Life-Cycle of U.S. Businesses”, , 20151, URL `https://ideas.repec.org/p/nbr/nberwo/20151.html` (2014).

Eaton, J. and S. Kortum, “Technology, Geography, and Trade”, Econometrica **70**, 5, 1741–1779, URL `https://ideas.repec.org/a/ecm/emetrp/v70y2002i5p1741-1779.html` (2002).

Ebenstein, A., “The consequences of industrialization: Evidence from water pollution and digestive cancers in china”, The Review of Economics and Statistics **94**, 1, 186–201 (2012).

Economy, E. C., *The River Runs Black: The Environmental Challenge to China's Future* (Cornell University Press, 2004).

Foster, L., J. Haltiwanger and C. J. Krizan, "Market Selection, Reallocation, and Restructuring in the U.S. Retail Trade Sector in the 1990s", The Review of Economics and Statistics **88**, 4, 748–758, URL `https://ideas.repec.org/a/tpr/restat/v88y2006i4p748-758.html` (2006).

Goulder, L. H., "Environmental taxation and the "double dividend": A reader's guide", NBER Working Paper No. 4896 (1994).

Graff Zivin, J. and M. Neidell, "Environment, health, and human capital", Journal of Economic Literature **51**, 3, 689–730, URL `http://www.aeaweb.org/articles.php?doi=10.1257/jel.51.3.689` (2013).

Grossman, G. M. and A. B. Krueger, "Environmental impacts of a north american free trade agreement", in "The U.S.-Mexico Free Trade Agreement", edited by P. M. Garber (MIT Press, Cambridge, Massachusetts, 1993).

Grossman, G. M. and A. B. Krueger, "Economic growth and the environment", The Quarterly Journal of Economics **110**, 2, 353–377, URL `http://www.jstor.org/stable/2118443` (1995).

Grossman, G. M. and E. Rossi-Hansberg, "Trading Tasks: A Simple Theory of Offshoring", American Economic Review **98**, 5, 1978–97, URL `https://ideas.repec.org/a/aea/aecrev/v98y2008i5p1978-97.html` (2008).

Guner, N., A. Parkhomenko and G. Ventura, "Managers and Productivity Differences", , 9586, URL `https://ideas.repec.org/p/iza/izadps/dp9586.html` (2015).

Guner, N., G. Ventura and Y. Xu, "Macroeconomic implications of size-dependent policies", Review of Economic Dynamics **11**, 4, 721–744 (2008).

Hall, R. E. and C. I. Jones, "Why do some countries produce so much more output per worker than others?", The Quarterly Journal of Economics **114**, 1, 83–116 (1999).

Head, K. and J. Ries, "FDI as an outcome of the market for corporate control: Theory and evidence", Journal of International Economics **74**, 1, 2–20, URL `https://ideas.repec.org/a/eee/inecon/v74y2008i1p2-20.html` (2008).

Hopenhayn, H. A., "Entry, exit, and firm dynamics in long run equilibrium", Econometrica **60**, 5, 1127–1150, URL `http://www.jstor.org/stable/2951541` (1992).

Hopenhayn, H. A., "On the measure of distortions", NBER Working Paper No. 20404 (2014).

Hopenhayn, H. A. and R. Rogerson, "Job turnover and policy evaluation: A general equilibrium analysis", Journal of Political Economy **101**, 5, 915–938, URL `http://www.jstor.org/stable/2138602` (1993).

Hsieh, C.-T. and P. J. Klenow, "Misallocation and manufacturing TFP in China and India", The Quarterly Journal of Economics **124**, 4, 1403–1448, URL `http://qje.oxfordjournals.org/content/124/4/1403.abstract` (2009).

Hsieh, C.-T. and P. J. Klenow, "The life-cycle of plants in India and Mexico", The Quarterly Journal of Economics **129**, 3, 1035–1084 (2014).

Hsieh, C.-T. and B. A. Olken, "The missing "missing middle"", Journal of Economic Perspectives **28**, 3, 89–108 (2014).

Jia, R., "Pollution for promotion", Manuscript (2014).

Jiang, L., C. Lin and P. Lin, "The determinants of pollution levels: Firm-level evidence from chinese manufacturing", Journal of Comparative Economics **42**, 1, 118–142 (2014).

Kapika, M., "How Important Is Technology Capital for the United States?", American Economic Journal: Macroeconomics **4**, 2, 218–48, URL `https://ideas.repec.org/a/aea/aejmac/v4y2012i2p218-48.html` (2012).

Kuznets, S., "Economic growth and income inequality", The American Economic Review **45**, 1, 1–28 (1955).

Lagakos, D., "Explaining cross-country productivity differences in retail trade", Journal of Political Economy, forthcoming (2016).

Lin, J., D. Pan, S. J. Davis, Q. Zhang, K. He, C. Wang, D. G. Streets, D. J. Wuebbles and D. Guan, "Chinas international trade and air pollution in the united states.", Proceedings of the National Academy of Sciences **111**, 5, 1736–1741 (2014).

Lucas, R. E. J., "On the size distribution of business firms", The Bell Journal of Economics **9**, 2, 508–523, URL `http://www.jstor.org/stable/3003596` (1978).

Luttmer, E. G. J., "Selection, growth and the size distribution of firms", The Quarterly Journal of Economics **122**, 3, 1103–1144 (2007).

Markusen, J. R., "Multinationals, multi-plant economies, and the gains from trade", Journal of International Economics **16**, 3-4, 205–226, URL `https://ideas.repec.org/a/eee/inecon/v16y1984i3-4p205-226.html` (1984).

McGrattan, E. R. and E. C. Prescott, "Openness, technology capital, and development", Journal of Economic Theory **144**, 6, 2454–2476, URL `https://ideas.repec.org/a/eee/jetheo/v144y2009i6p2454-2476.html` (2009).

McGrattan, E. R. and E. C. Prescott, "Technology Capital and the US Current Account", American Economic Review **100**, 4, 1493–1522, URL `https://ideas.repec.org/a/aea/aecrev/v100y2010i4p1493-1522.html` (2010).

Melitz, M. J., "The impact of trade on intra-industry reallocations and aggregate industry productivity", Econometrica **71**, 6, 1695–1725, URL `http://www.jstor.org/stable/1555536` (2003).

Melitz, M. J. and G. I. P. Ottaviano, "Market size, trade and productivity", Review of Economic Studies **75**, 1, 295–316 (2008).

Olley, G. S. and A. Pakes, "The dynamics of productivity in the telecommunications equipment industry", Econometrica **64**, 6, 1263–1297 (1996).

Olmstead, A. L. and P. W. Rhode, "Reshaping the landscape: The impact and diffusion of the tractor in american agriculture, 1910—1960", The Journal of Economic History **61**, 3, 663–698, URL `http://www.jstor.org/stable/2698132` (2001).

Parente, S. L. and E. C. Prescott, "Barriers to technology adoption and development", The Journal of Political Economy **102**, 2, 298–321 (1994).

Poschke, M., "The firm size distribution across countries and skill-biased change in entrepreneurial technology", Manuscript (2015).

Ramondo, N., "A quantitative approach to multinational production", Journal of International Economics **93**, 1, 108–122, URL `https://ideas.repec.org/a/eee/inecon/v93y2014i1p108-122.html` (2014).

Restuccia, D. and R. Rogerson, "Policy distortions and aggregate productivity with heterogeneous establishments", Review of Economic Dynamics **11**, 4, 707–720 (2008).

Restuccia, D. and R. Rogerson, "Misallocation and productivity", Review of Economic Dynamics **16**, 1, 1–10 (2013).

Rossi-Hansberg, E. and M. L. J. Wright, "Establishment size dynamics in the aggregate economy", American Economic Review **97**, 5, 1639–1666 (2007).

Schankerman, M. and A. Pakes, "Estimates of the Value of Patent Rights in European Countries during the Post-1950 Period", Economic Journal **96**, 384, 1052–76, URL `https://ideas.repec.org/a/ecj/econjl/v96y1986i384p1052-76.html` (1986).

Shadbegian, R. J. and W. B. Gray, "Pollution abatement expenditures and plant-level productivity: A production function approach", Ecological Economics **54**, 2, 196–208 (2005).

Shapiro, J. S. and R. Walker, "Why is pollution from U.S. manufacturing declining? the roles of trade, regulation, productivity and preferences", NBER Working Paper No. 20879 (2015).

Song, Z., K. Storesletten and F. Zilibotti, "Growing like China", American Economic Review **101**, 1, 196–233 (2011).

Song, Z. and G. Wu, "Identifying capital misallocation", Manuscript (2015).

Tombe, T. and X. Zhu, "Trade, migration and productivity: A quantitative analysis of china", Manuscript (2015).

Vennemo, H., K. Aunan, H. Lindhjem and H. M. Seip, "Environmental pollution in china: Status and trends.", Review of Environmental Economics and Policy **3**, 2, 209–230, URL https://libproxy.cc.stonybrook.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eih&AN=44545982&site=ehost-live&scope=site (2009).

Veracierto, M., "Employment Flows, Capital Mobility, and Policy Analysis", International Economic Review **42**, 3, 571–95, URL https://ideas.repec.org/a/ier/iecrev/v42y2001i3p571-95.html (2001).

Wang, J. and J. Whalley, "Are chinese markets for manufactured products more competitive than in the us? a comparison of china-us industrial concentration ratios", NBER Working Paper 19898 (2014).

Wang, M., M. Webber, B. Finlayson and J. Barnett, "Rural industries and water pollution in china", Journal of Environmental Management **86**, 4, 648–659 (2008).

World Bank, *China Air, Land and Water: Environmental Priorities for a New Millenium* (World Bank Publications, Washington DC, 2001).

Yang, G. and D. Zhuang, *Atlas of the Huai River Basin Water Environment: Digestive Cancer Mortality* (Springer, New York, 2014).

Young, A., "The razor's edge: Distortions and incremental reform in the people's republic of china", The Quarterly Journal of Economics **115**, 4, 1091–1135 (2000).

Zhang, J., "The impact of water quality on health: Evidence from the drinking water infrastructure program in rural china", Journal of Health Economics **31**, 1, 122–134, URL http://www.sciencedirect.com/science/article/pii/S0167629611001172 (2012).

Zheng, S. and M. E. Kahn, "Understanding china's urban pollution dynamics", Journal of Economic Literature **51**, 3, 731–772, URL http://www.aeaweb.org/articles.php?doi=10.1257/jel.51.3.731 (2013).

APPENDIX A

ACCOUNTING EXERCISES FOR CHAPTER 2

This appendix provides robustness calculations of the accounting exercises.

Estimation Strategies.—Ideally, we would like to have the pollution intensity over firm size. However, such data do not exist since the NGSPS only reports total value of production and total amount of pollution. Therefore, we would need to construct pollution intensity over the number of employees. We use the CNEC for this purpose. In particular, we estimate the corresponding bins for production for each employment bin from the U.S data.

SUSB reports firm size in 22 bins. For the U.S size bins, we construct the corresponding production bins to be used in NGSPS. CNES is used to bridge the employment bins (SUSB) to the production bins (NGSPS).

1. *Non-parametric*:
    - For each U.S employment bin, we compute the 1st quartile and 3rd quartile production levels for Chinese firms within that employment bin. The two quartiles are used as the lower and upper bounds for the production bins in NGSPS.
    - We then use the median pollution intensity of firms within the newly defined production bins as the average pollution intensity for those bins.
    - Lastly, we calculate the aggregate pollution by assigning to each bin the corresponding share of production. The NGSPS production bins are used for China and the employment bins are used for U.S.
2. *Piecewise Linear*:
    - For each U.S employment bin, we regress log-product on log-employment using the subset of Chinese firms within that employment bin. The lower and upper bounds for the production bins in this case are calculated as the predicted value of the above regression.
    - We then run piecewise log-linear regression of pollution intensity on production within each new production bin. The average pollution intensity is chosen to be the predicted intensity at the midpoint of the new log-production bin.
    - Lastly, the average intensity is applied to the production share distributions. The U.S distribution does not change, however, a new distribution for China is calculated since the endpoints of the production bins are different.
3. *Parametric*:
    - Using CNEC, we regress log-production on log-number of workers, which yields a parametric relationship between the number of workers and production.
    - Using NGSPS, we regress log-intensity on log-production, which yields a parametric relationship between intensity and production. From these two relationships, we can subsequently construct a new *parametric* relationship

Methods	Paper	Agri	Tex	Chem	Bever	Avg	Reduc
Non-parametric	39.8%	60.7%	81.6%	102.5%	103.7%	63.5%	28.2%
Piecewise-linear	34.8%	69.4%	93.5%	180.1%	N/A[a]	75.4%	19.0%
Parametric	43.5%	61.1%	97.5%	101.2%	89.0%	67.0%	25.5%

† Note: Please see notes of Table 2.2 for acronyms of industries. For individual industries, the numbers reported are the aggregate pollution from the artificial U.S production structure as percentage from that of China. We use the 1st and 3rd quartile in the non-parametric calculation. Column 6 (Ave) calculates the weighted average of these ratios using the percentage contribution in row one of Table 2.2 as weights. Column 7 (Reduc) reports the aggregate reduction, which is simply the average without normalization.

[a] Since the beverage industry has a lot fewer firms than the others, there are employment size bins with no corresponding firms in China, which invalidates the method. We set the ratio to 100% in the calculation of the last two averages.

Table A.1: Size Distribution on Pollution

between intensity and number of employees. The average intensity is chosen to be the midpoint of each U.S employment bin. Notice that in this case we have a direct functional form for employment and intensity).

- Lastly, the average intensity is applied to the production share distributions. The U.S distribution does not change, but a new distribution for China is calculated since the endpoints of the production bins are different. Notice that this distribution for China is the one we use in Section I.C.

The estimation results are shown in Table A.1. Each of these three methods has its own advantages and disadvantages. The two non-parametric methods capture more of the variation at the local level, which could be washed out in a parametric estimation across the whole state space. However, this local nature also introduces a lot of instability on the estimations. Further, there are situations when there are gaps not covered by adjacent production bins and situations when these production bins overlap with each other. Under these conditions, some information will be lost while other is used for multiple times. Nevertheless, the results are robust across different estimation strategies.

APPENDIX B

FORMAL PROOFS TO THE MAIN RESULTS IN SECTION 2 OF CHAPTER 2

In this section we provide formal proofs to the results in Section II.C. For convenience, we state those results here again.
PROOF OF LEMMA 2:

Lemma 2. *In an economy with no product market frictions, $\pi_0(z)$ and $\pi_1(z)$ are both increasing and linear with respect to z. In addition, the slope of $\pi_1(z)$ is steeper than that of $\pi_0(z)$:*

$$\frac{\partial \pi_0(z)}{\partial z} = (1-p\xi)\frac{\partial \pi_1(z)}{\partial z}, \qquad \forall z \in Z \tag{B.1}$$

Proof. Since k_E is sunk-cost, it does not affect firms' decision once is paid. The factor demand decisions for the two types of firms are therefore the same. The first order conditions for capital and labor are respectively

$$\frac{\partial \pi_i(z)}{\partial k}: \qquad \alpha\gamma z^{1-\gamma}k^{\alpha\gamma-1}n^{(1-\alpha)\gamma} = R \tag{B.2}$$

$$\frac{\partial \pi_i(z)}{\partial n}: \qquad (1-\alpha)\gamma z^{1-\gamma}k^{\alpha\gamma}n^{(1-\alpha)\gamma-1} = W, \qquad i = 0,1 \tag{B.3}$$

Dividing (B.2) with (B.3) yields constant capital to labor ratio h

$$h = \frac{k}{n} = \frac{\alpha W}{(1-\alpha)R} \tag{B.4}$$

which says more capital is demanded when technology is capital intensive (higher α) or when capital rental price R low. Notice that the system of equations (B.2) with (B.3) is log-linear and thus has closed-form solution. With some algebra, the solutions are characterized by

$$n(z) = \Phi_1 R^{\frac{\alpha\gamma}{\gamma-1}} W^{\frac{1-\alpha\gamma}{\gamma-1}} \cdot z, \qquad \Phi_1 = \left[\frac{(1-\alpha)^{\alpha\gamma}}{(1-\alpha)\gamma\alpha^{\alpha\gamma}}\right]^{\frac{1}{\gamma-1}} \tag{B.5}$$

$$k(z) = \Phi_2 R^{\frac{1+\gamma(\alpha-1)}{\gamma-1}} W^{\frac{\gamma(1-\alpha)}{\gamma-1}} \cdot z, \qquad \Phi_2 = \frac{\alpha}{1-\alpha}\Phi_1 \tag{B.6}$$

Substitute the optimal solutions (B.5) and (B.6) back to the definition of profits functions (2.8) and (2.9), we have

$$\pi_0(z) = (1-p\xi)\left(\Omega - \frac{1}{1-\alpha}\Phi_1\right)\kappa z, \qquad \Omega = \left(\frac{\alpha}{1-\alpha}\right)^{\alpha\gamma}\Phi_1^{\gamma} \text{ and } \kappa = W^{\frac{\gamma(1-\alpha)}{\gamma-1}} R^{\frac{\alpha\gamma}{\gamma-1}}$$

$$\pi_1(z) = \left(\Omega - \frac{1}{1-\alpha}\Phi_1\right)\kappa z - Rk_E$$

where it is clear that both functions are increasing and linear in z and (B.1) is true. □

PROOF OF COROLLARY 2:

Corollary 2. *Suppose the product market frictions are specified as* $\max\left\{0, 1 - z^{\phi_1}\right\}$ *with* $1-\gamma+\phi_1 > 0$*, then* $\pi_0(z)$ *and* $\pi_1(z)$ *are both increasing and concave with respect to* z*. In addition, the slope of* $\pi_1(z)$ *is steeper than that of* $\pi_0(z)$*:*

$$\frac{\partial \pi_0(z)}{\partial z} = (1 - p\xi)\frac{\partial \pi_1(z)}{\partial z}, \qquad \forall z \in Z$$

Proof. The proof is straightforward given Lemma 2. Substituting in the tax function, $\pi_0(z)$ and $\pi_1(z)$ now becomes

$$\pi_0(z) = (1 - p\xi)\left(\Omega - \frac{1}{1-\alpha}\Phi_1\right)\kappa z^{\frac{1-\gamma+\phi_1}{1-\gamma}}$$
$$\pi_1(z) = \left(\Omega - \frac{1}{1-\alpha}\Phi_1\right)\kappa z^{\frac{1-\gamma+\phi_1}{1-\gamma}} - Rk_E$$

where Ω, Φ_1 and κ are defined as in Lemma 2.

Assumption $1 - \gamma + \phi_1 > 0$ guarantees the monotonicity of the profits functions. Concavity is easily verified by taking second order derivatives. □

PROOF OF PROPOSITION 3:

Proposition 3. *There exists a unique threshold* $\hat{z}$ *such that all household members with* $z \leq \hat{z}$ *choose to be workers and those with* $z \geq \hat{z}$ *become entrepreneurs. Further,* $\hat{z}$ *is the solution of* $W = \pi(\hat{z})$

Proof. Since the overall profit function $\pi(z)$ is the upper envelope of $\pi_0(z)$ and $\pi_1(z)$, from Lemma 2 and 2 we know $\pi(z)$ is monotonic increasing. It is easy to verify that $\pi(0) = 0$. Therefore, as long as $0 < W < \pi(\overline{z})$, we can find a unique $\hat{z}$ such that $\pi(\hat{z}) = W$, where uniqueness follows from monotonicity. The condition $0 < W < \pi(\overline{z})$ is guaranteed in the general equilibrium version of our model by Inada condition on the production function. □

PROOF OF PROPOSITION 4:

Proposition 4. *Given* k_E, W, τ_z *and* R*, there exist unique thresholds* $\tilde{z}_n$ *and* $\tilde{z}_f$ *such that:*

(i) In the economy with no product market frictions, entrepreneurs with $z \leq \tilde{z}_n$ *produce using dirty technology while those with* $z > \tilde{z}_n$ *produce using clean technology.*

(ii) In the economy with product market frictions, entrepreneurs with $z \leq \tilde{z}_f$ *produce using dirty technology while those with* $z > \tilde{z}_f$ *produce using clean technology.*

(iii) $\tilde{z}_n < \tilde{z}_f$*, that is, product market frictions impede the technology upgrade.*

Proof. Uniqueness follows from

$$\frac{\partial \pi_0(z)}{\partial z} = (1 - p\xi)\frac{\partial \pi_1(z)}{\partial z}, \qquad \forall z \in Z$$

and monotonicity under both the case with and without frictions.

We can solve for analytical expression for z_n:

$$z_n = \frac{Rk_E}{p\xi\left[\Omega\kappa^\gamma - \frac{1}{1-\alpha}\kappa\right]} \tag{B.7}$$

where z_n is simply the "distance" (Rk_E) over "speed" $\left(p\phi\left[\Omega\kappa^\gamma - \frac{1}{1-\alpha}\kappa\right]\right)$. The "distance" in both cases are the same, so eventually whether z_f lies left or right to z_n depends on the "speed" of convergence.

Using expressions of the profits functions with frictions, we can show that

$$z_n = \frac{Rk_E}{p\xi\left[\Omega - \frac{1}{1-\alpha}\right]\kappa} < \left(\frac{Rk_E}{p\xi\left[\Omega - \frac{1}{1-\alpha}\right]\kappa}\right)^{\frac{1-\gamma}{1-\gamma+\phi_1}} = z_f^* \tag{B.8}$$

which proves the proposition.

One caveat is that the second inequality holds only if the number in the parentheses is greater than 1. We verify this in our quantitative analysis but restrain ourselves from discussing extreme cases where the condition is not hold. □

APPENDIX C

THE COSTS OF ABATEMENT TECHNOLOGY ADOPTION FOR CHAPTER 2

In the paper, we model the costs of adopting abatement technology as a fixed costs k_E, regardless of the production scale of firms. The readers may concern that this could be a mis-specification. Two alternatives arise naturally: fixed costs plus operating costs and scale-related fixed costs. In this section, we show that each alternative contradicts some of the empirical evidence we document from the data.

Operating Costs.—We start with the case of operating costs. Trimming the top 1% observations for outliers, the distribution of the ratio of operating costs over total value of production is shown in Table C.1. On average, operating costs of abatement

Minimum	25%	50%	75%	Maximum	Mean
0	0.0013	0.0049	0.0143	0.2084	0.0150

Table C.1: Operating Costs as a Fraction of Output

equipment takes about 1.5% of a firm's annual value of production. In addition, the median of the ratio is less than 0.5%, suggesting that operating costs are negligible for more than 50% of the firms. For conciseness consideration, we choose to omit them from the model.

Fixed Costs Proportional to Production Level.—We do observe in the data that firms with larger production scale tend to make larger investment in abatement equipments. For instance, the correlation between the log value of production and abatement equipment investment is equal to 0.64. However, the relationship significantly weakens if we remove the percentage interpretation implied by taking logarithmic. In fact, if we calculate the ratio of abatement equipment investment over total production of firms below certain quantiles, the ratio decreases gradually as we include more large firms into the calculation. Table C.2 reports the ratio of *aggregate* treatment equipments installation costs as a fraction of *aggregate* value of output when firms with size below certain quantiles are included. Recall that we have documented that

25%	40%	60%	80%	95%
0.6195	0.2475	0.1151	0.0668	0.0433

Table C.2: Installation Costs as a Fraction of Output

the unit-cost per processing capacity is also decreasing with the equipment capacity, together these facts suggest the existence of returns to scale of treatment equipments.

Relative Prices of Physical versus Biological Equipment.—We assume in the main text that only the installation of biological equipments requires a fixed costs. In this paragraph, we provide evidence in support of this choice. The distributions of prices for physical and biological equipments in absolute term (CNY 10,000 in year 2007) are listed in Table C.3. As is shown in Table C.3, the costs of biological equipments is 7 to 15 times of those of the physical equipments. We further calculate the ratio of *aggregate* treatment equipments installation costs as a fraction of *aggregate* value of output for different technologies. We find that the ratio is 3.07% for physical technology and 12.83% for biological technology. In addition, the total investment of physical equipment over biological equipment is 0.087, meaning that the installation

Technologies	Minimum	25%	50%	75%	Maximum	Mean
Physical	0.002	2.000	6.000	25.000	800.000	36.140
Biological	0.04	30.00	85.00	260.00	4000.00	249.90

Table C.3: Relative Prices of Treatment Equipments

costs of physical equipments when aggregated across the economy are only 9% of those of the biological equipments.

These evidence suggest that comparing to biological equipment, the costs of installing physical treatment equipments are nearly negligible. Therefore in our later calculation we assume that only the adoption of biological equipment is costly.

In summary, the evidence suggests that firms' investment in abatement equipment is increasing with production scale but to a much lesser extent in absolute term. Again for simplicity and without loss of generality, we choose the fixed costs modeling strategy.

BIOGRAPHICAL SKETCH

Xican Xi is from Ningbo, China. He earned a bachelor's degree in Economics from the School of Economics, Zhejiang University in 2007 and a master's degree in Finance from the School of Economics, Zhejiang University in 2010.

During his stay in the graduate school at the Arizona State University, Xican served as a Research Assistant to Professors Edward C. Prescott and Todd Schoellman. He also served as a Research Analyst at the Federal Reserve Bank of Minneapolis during the summer of 2012. He was honored to receive the Hardison Award for the best microeconomics and macroeconomics comprehensive examinations, the Prescott Award for summer research support, and Barchilon Award for best progress towards dissertation. He also earned a master's degree in Economics from the university in 2012.

From the Summer of 2016 to the Spring of 2017, Xican will serve as an Economist in International Monetary Fund at Washington D.C. Starting in the Spring of 2017, he will serve as an Assistant Professor of Economics at Fudan University in China.